

Vol. 37, No. 2

February, 1940

Psychological Bulletin

A CRITICAL EXAMINATION OF THE UNIVERSITY OF IOWA STUDIES OF ENVIRONMENTAL INFLUENCES UPON THE IQ

BY QUINN McNEMAR
Stanford University

INTRODUCTION

During the last year or so the educational journals of the country have contained many articles from the pen of either Professor George D. Stoddard or Dr. Beth L. Wellman on changes in the IQ brought about by school and environmental factors. This popularization of research results has not been confined to learned journals, but has been spread abroad *via* more popular routes such as the newspapers and the radio. In an article in the *New York Times* under the signature of Dr. Wellman (28) will be found such statements as: "The extent of change under especially favorable circumstances may be sufficient to move a child from average intelligence to the so-called genius or extremely high levels. Or it may when conditions are especially unfavorable change children from average intelligence to feeble-mindedness." In *Childhood Education* (27) Dr. Wellman writes that the "group mental level of the children in a school is an important factor in the change in IQ of a particular child." In the *Journal of Consulting Psychology* Dr. Wellman says that "results from long-time consecutive studies of intelligence of children are demanding certain changes in our concept of intelligence in order that our concept conform with the facts. Data showing large changes in IQ have been steadily piling up, until they can no longer be summarily waved aside. There is no escape from the fact that the IQ's of children have possibilities of change over a large portion of the IQ range from genius to feeble-mindedness" (25, p. 97).

In an article under the co-authorship of Dr. Wellman and Professor Stoddard in *Social Frontier* (31) it is said that "some geniuses

are made" and "some children are made feeble-minded" by their environments. Professor Stoddard, in *National Parent-Teacher* (17), after discussing marked changes in IQ says: "Such changes are not to be considered artificial or transitory." "They are durable." In a lecture on "The IQ: its ups and downs," by Professor Stoddard, one finds the statement: "The children of definitely moronic mothers and laboring class fathers, if placed early in good foster homes, will turn out to be above average in mental ability" (18, p. 49).

In *Progressive Education* Professor Stoddard (14) writes that "even scientific rigor and caution need not prevent us from saying flatly" that better nursery schools foster mental development. In the *Proceedings of the National Education Association* (15) Stoddard goes on record to the effect that "some of our recent well-documented work at Iowa . . . indicate[s] definitely that the intelligence quotients of young children can be raised by environmental stimulation." A recent issue of *School & Society* (16) contains an address of Professor Stoddard's in which he speaks of "the reaffirmation, in a most technical and substantial way, of the idea that the child is plastic," and states that "the scientific evidence against such a stand [that plasticity does not apply to the intelligence quotient] is mounting and cannot be denied."

That the new gospel is being carried beyond the educational journals is again evidenced by Dr. Wellman's article in the *Journal of Home Economics* (30), in which will be found the usual citation of changes for selected cases and the extraordinary statement that the orphanage children not enrolled in preschool "moved swiftly in the direction of feeble-mindedness." We thus see from all of these quotations not only how positive are the generalizations from the Iowa studies, but also how widely the news is being spread.

The average reader will naturally assume that claims so extreme would surely not be made by well-known psychologists without the best of evidence. That the claims may sound exaggerated is anticipated in a statement of Wellman, Skeels, et al.: "These statements may seem unbelievably extreme" (10, p. 185). The writer does not believe that either the environmentalist or the hereditarian can be blamed for expressing a word of skepticism, but he believes that skepticism should lead to a minute examination of the research findings instead of resulting merely in sweeping condemnations thereof. In the critique to follow, we shall first examine in some detail the orphanage preschool project of Skeels, Updegraff, Wellman, and Williams, and then scrutinize the foster children study of Skodak

and Skeels. Other studies will then be briefly considered, and finally we shall devote a section to a type of statistical treatment which is common to all the IQ studies of the Iowa Child Welfare Research Station.

THE ORPHANAGE PRESCHOOL PROJECT

"A study of environmental stimulation" is the title of a monograph under the co-authorship of Skeels, Updegraff, Wellman, and Williams (10). On the grounds of an orphanage, which is described as a nonstimulating atmosphere intellectually, a model preschool was established. For every child, age 18 months to $5\frac{1}{2}$ years, who attended this preschool there was a control child whose activities were only those of the restricted orphanage environment, and who had been paired with the preschool child on the basis of CA, MA, IQ, sex, nutritional status, and length of residence in the orphanage. There were 21 such pairs of children at the beginning of the project, but, due to the placement of children in foster homes, only a few of these continued for the full three years of the study. As children dropped out of either the preschool or control groups, new admissions were added with an attempt continually to equalize the two groups in relation to the factors considered in the original matching. Data are given to show that the original groups were equal with regard to CA, MA, and IQ, but the information given for the total groups, 46 preschool and 44 controls, indicates that the preschool group averaged 4.5 IQ points higher on the initial tests.

Initial IQ's were available for both the preschool and control individuals prior to school experience by the former. Over a period of three years retests were given at intervals of "approximately six months." Since, save for the five or six hours of daily school attendance of the one group, the two groups experienced the same living conditions in a nonstimulating orphanage environment, it would seem that any found differences in the mental development of the two groups could be attributed to school attendance. However, that the setup is not nearly so ideal as one might at first suppose is evidenced by the fact that, for any given comparison regarding the effect of preschool experience on IQ's, it was not possible to control more than two or three of the following variables at a time: (1) age, (2) initial intelligence, (3) orphanage residence, (4) actual number of days of school attendance, (5) days of residence between tests or retest intervals, (6) various examiners, (7) practice effects, (8) Kuhlmann-Binet or Stanford-Binet, (9) possible unintentional coaching in pre-

school on material similar to many items in the tests used, and (10) differences in rapport in testing.

The first, and perhaps chief, finding regarding changes in intelligence is summarized herewith in Table I, which is a condensation of the authors' Table 2. The reader will readily see that no allowance has been made here for certain variables mentioned above (though the authors in their complete table do make some allowance for initial IQ by separate treatment for those initially above and below

TABLE I
IQ CHANGES ACCORDING TO DAYS RESIDENCE BETWEEN TESTS

Residence	Preschool		Control		Difference in Change	D/σ_D
	N	Mean IQ	N	Mean IQ		
1 to 199 days	91	87.6	76	82.4		
	91	86.9	76	82.6	.2	.9
		—.7		.2		.7
200 to 399 days	90	82.3	96	77.7		
	90	86.0	96	76.5	-1.2	-4.9
		3.7		-1.2		3.3
400 or more days	40	80.1	65	77.2		
	40	84.7	65	72.6	-4.6	-9.2
		4.6		-4.6		4.2

80 IQ). The excess of the N's in this table over 46 and 44 is due to the fact that "a child was included as a separate case as many times as he met the requirement of residence interval" (days between tests). Thus, a child who had four tests was included in the appropriate residence group for first to second, second to third, third to fourth, first to third, first to fourth, and second to fourth tests. In other words, a child with n tests contributes $n(n-1)/2$ to the N's of Table I. This is the first of a series of jugglings in a monograph which is literally filled with highly questionable procedures. In the first place, one can raise a question as to the meaning that can be attached to an average change so determined, but, granting that it does have a definite meaning, one must ask about the proper formula for evaluating the changes in terms of sampling. If the authors care to dig deep enough into the statistics of sampling, they will find that the fundamental conditions for sampling are that each unit or individual in the universe being sampled must have an equal chance of being included in the sample, and that the drawing of one unit or individual must in no way affect the drawing of any other unit or individual. This latter condition can be stated differently: The drawing of each unit or individual must be independent of the drawing of any other unit or individual. These two conditions are the

basic assumptions for all ordinary standard error formulas, and any departure from either of these two fundamental conditions invalidates the use of such formulas.

That the authors have violated the second condition of sampling is so obvious that we should not have to elaborate thereon. Once a child has been included, he is automatically included two more times if he has had three tests, or five more times if he has had four tests, or nine more times if five tests have been administered. Nevertheless, he is just one out of a defined universe of individuals who might be drawn for the sample. To say that this type of thing has operated similarly for the preschool and control groups, and therefore has not affected the results, is to miss the point.

In the absence of any adequate sampling error formulas for testing significance where we have this double type of sampling—a sample of individuals and a varying number of observations as we pass from individual to individual—we must either manufacture a formula for the situation or alter the situation in such a way as to permit the proper use of available formulas. Mathematical statisticians have found that formulas for similar situations are not easily derived unless certain simplifying assumptions are permissible which, when made, so restrict the obtained formulas as to make them inapplicable to the practical situation.¹ How, then, can we handle such data as these authors have collected? A reasonable way for surmounting the difficulty would be to state the problem as follows: Given two groups with known (measured) IQ's at the beginning of a project, one group being subject to stimulating nursery school conditions, the other group (control) remaining submerged in an intellectually stifling orphanage, then we can easily check the effect of preschool attendance on mental development providing we are willing to grant two assumptions: (1) that the initial IQ's can be taken as representing the intellectual status of the two groups at the beginning of the experiment, and (2) that the last, or final, IQ's can be taken as reflecting the later intellectual standing of the groups. A straightforward comparison of the mean changes from initial to final IQ's will permit conclusions as to the difference in gain or loss for the two groups. The sampling evaluation of found differences between the groups will involve nothing more complicated than the standard errors of two mean changes and avoids the indefensible procedure of using inflated N's.

¹ See Fisher, R. A. The statistical utilizations of multiple measurements. *Ann. Eugen., Camb.*, 1938, 8, 376-386.

For those who object that such a simplifying procedure does not take into account progressive changes, let it merely be noted that, after it has been demonstrated that a real difference in change between the groups has taken place, it will not be too late to examine the progress of the changes.

From the original data, which have been kindly supplied by Dr. Wellman, we have determined the mean changes from initial to final test for three residence intervals. In doing this we have not attempted to make any allowance for the aforementioned uncontrolled variables. The results of this treatment of changes, summarized in Table II,

TABLE II
IQ CHANGES BETWEEN INITIAL AND FINAL TESTS ACCORDING TO RESIDENCE
BETWEEN INITIAL AND FINAL TESTS

Residence	Preschool		Control		Difference in Change	D/σ_D
	N	Mean IQ	N	Mean IQ		
1 to 199 days	15	98.5	11	90.8		
	15	96.9	11	94.6	-1.6	1.2
200 to 399 days	10	79.4	11	80.4		
	10	84.5	11	82.8	5.1	.9
400 or more days	21	82.9	22	81.4		
	21	85.7	22	75.7	2.8	2.2
					-5.7	-8.5

should be compared with the authors' findings as given in our Table I. In view of the fact that a proper statistical treatment of the data has reduced their two significant critical ratios of 3.3 and 4.2 to .9 and 2.2, we are inclined to say that the authors' first, and main, finding has resulted from faulty statistics rather than from preschool attendance. Incidentally, the use of inflated N's not only exaggerates the statistical significance of the findings, but is also apt to mislead the layman into placing undue confidence in a result. For instance, in one of Dr. Wellman's popular articles we find the passage: "In two years' time a group of twenty-six children who averaged 90 in IQ dropped 16 points" (27). The correct N happens, in this case, to be only 11. It will be noted that such changes as have taken place, though of doubtful statistical significance, are in the direction of gains for the preschool and losses for the control group, but before we can attribute such changes to the stimulating effect of the preschool environment and the progressively degrading influence of the orphanage, we must ask about other factors which might account for the changes.

One of the most important of the uncontrolled variables in the study has to do with the rapport between child and examiner. Nothing is said in the report about shyness, negativism, distractibility, or general coöperativeness of the children during the testing. The report does contain some information (pp. 23-25) which is highly pertinent. In their picturesque description of the children and their habits and attitudes which prevailed at the beginning of the project the authors make such assertions as the following: "The language of the children was in the great majority of cases either entirely or practically unintelligible"; "any constructive conversation seemed out of the question"; the children were "not accustomed to listening to the words of adults"; "the attitude toward adults was a strange mixture of defiance, wish for affection, and desire for attention"; there existed "a feeling of the individual against the world, expecting no quarter and giving none"; reaction to strangers was "the same [as] to wax figures"; there was "an almost invariable negative response to anything which the child could possibly interpret as potential coercion" and a "highly emotional response to unwelcome requests"; the children were "full of suspicion and mistrust" and "seldom in the frame of emotion or mind to face a situation"; they showed "lack of confidence in adults" and "generally violent and moblike reactions to new situations."

If this description is not highly journalistic, then the initial tests were invalid to begin with and unworthy of further consideration. It might be argued that this lack of rapport would be alike for the initial tests on the two groups, but what about later tests? The controls remain in the "bad" orphanage and would therefore be expected to become even less coöoperative with the passing of time, whereas the preschool children would become more coöoperative as a result of decent treatment by adults in the preschool. This factor alone might serve as an adequate explanation (if a statistically insignificant difference calls for explanation) for such divergences as exist. It is difficult to understand just how the authors could overlook this long-recognized and exceedingly important matter of rapport. At this point one is made to wonder why they forget a previous paper by one of them, Updegraff, in which it was reported, with substantial data, that "it has been found that an intelligence quotient obtained just previous to a young child's first experience in preschool is not reliable" (20, p. 164).

Let us now turn to the second main finding with regard to environmental effects on intelligence. The changes for both groups were

analyzed according to initial IQ level: below and above 80, and by 10-point classifications. Inflated N's are again used in determining standard errors, with the result that all the critical ratios are spuriously high. Here we note a failure to appreciate the phenomenon of regression, which must be taken into account in this type of analysis. Without elaboration at this time on regression effects, which will be dealt with more fully later, we note a most amazing statement made by the authors with regard to the leveling effect of the orphanage environment on the control children: "Regardless of the original classification, all the groups headed for a final classification between 70 and 79 IQ. The effect of long residence for the control children was thus a leveling one, tending to bring all children to high grade feeble-mindedness or borderline classification" (p. 45).

This conclusion is based on a completely erroneous line of reasoning. The authors might have been expected to know that the only thing which kept their constructed curves from showing a greater leveling effect was the fact that the test-retest correlation was not zero. The interpretation of the authors can be exploded with a bang by asking for data on the variability of the group on the final tests as compared with that for the initial tests. This pertinent information is not included in the monograph. Direct computation (from the original data supplied by Dr. Wellman) shows a S.D. of 13.2 for the final test as opposed to 13.9 for the initial test. No doubt the authors will themselves be surprised to learn that the leveling effect was such as to reduce the S.D. by .7 of an IQ point, or to reduce the variance by less than 1%. Even for the long-residence group the S.D. is reduced only from 15.0 to 13.1—an unreliable drop. Evidently the leveling effect of the orphanage environment did not take place until the data had passed through the statistical laboratory.

In the subsequent analysis and discussion of the gains, the most important finding to emerge is a correlation of .28 between the actual number of days attendance by the preschool individuals and change in IQ's. The probable error of .04 attached to this r is evidently based on a combination N, partly individuals and partly observations; otherwise, the probable error for 46 individuals would be .09. How much of this obtained correlation reflects changes as due to "increased ability" and how much has resulted from increased rapport cannot be determined. Further analysis shows the changes in relation to percentage of attendance, and here (p. 50) we have the following remarkable finding: "Although between the 91 to 100 per cent group and the 61 to 70 per cent group there was only seven days' difference

in actual attendance, the difference in IQ change was 7.0 points." This should be too much even for the most ardent nurturite.

We come next to a striking graph (Fig. 8, p. 54) which tells only a part of the story concerning IQ changes for the control group, in that it portrays individual curves only for *decreases* in IQ. The casual reader is likely to be much impressed by this visual demonstration of the effect of unfavorable circumstances on the IQ, but the critical reader's first reaction will be to raise a question concerning possible cases which might show an increase. The data for the seven curves in this figure may be briefly summarized in terms of loss from initial to final IQ: 103 to 60, 98 to 61, 86 to 62, 83 to 60, 85 to 71, 80 to 70, and 79 to 69. Elsewhere (25, p. 99), Wellman states that "these cases represent the trend for the larger group of which they were members." We are about to see that this is a gross misstatement of fact. Since no significance can be attached to the losses of 10 points by the last two cases, we have left four cases which show marked losses of 43, 37, 24, and 23 points, and one case which shows a loss of 14 points. But an examination of the original data reveals two control children who showed gains of 27 (70 to 97) and 22 (61 to 83) points, and one who gained 14 points. We find, moreover, that of 15 individuals in the control group who had received only two tests, one showed a loss of 18 points and four showed gains of 16 or more points. Evidently, residence in the orphanage did not tend "to bring *all* children to high grade feeble-mindedness or borderline classification" (p. 45; *italics ours*). Further examination of the original scores leads to the discovery of only three in the total preschool group, as compared to seven in the control group, who showed gains from initial to final of more than 14 points. These three children gained 25, 19, and 17 points. Thus, the greatest individual gains took place not in the preschool but in the control group. This fact is not mentioned, nor can it be discovered from their published data.

Further perusal of the data supplied by Dr. Wellman discloses additional facts which are highly pertinent. Despite the fact that the groups had comparable mean ages, 41.9 and 41.4 months, at the time of the initial tests (p. 17), there were seven in the control, as opposed to three in the preschool, group who were tested prior to 20 months of age. Now it happens that the three controls showing the largest losses (43, 37, and 24 points) were tested initially at less than 20 months of age. If we count the number tested prior to 24 months, we find seven for the preschool group and eleven for the

control group, and five of these eleven are among the seven control children who provided the conclusion "that children of average ability may be made feeble-minded" (p. 57). Surely some consideration should have been given to the question of the comparability of measures of "intelligence" at age 18 months with such measures at age 4 before even suggesting such a conclusion.

In summarizing a section on the later development, we learn that "of the children later adjudged feeble-minded and transferred to the school for feeble-minded, 75 per cent were from the control group" (p. 60). This percentage, coupled with the claim that the preschool and control groups were "originally equated" on the basis of intelligence, sounds quite impressive until it is recalled that the IQ's only of those in the groups at the beginning of the project were equated. For some unexplained reason the "attempt continually to equalize the two groups" when new admissions were added was not very successful. The 21 controls added five or more months after the beginning of the project averaged 83.3, as contrasted with 90.6 for the 23 new admissions to the preschool group. Another reason why the above percentage loses significance for all but casual summary-readers is the fact that it is based on only eight cases, six from the control and two from the preschool group.

The remaining part of the monograph deals principally with Merrill-Palmer tests, language development, general information, social maturity, and motor achievement. It is not our purpose to discuss these additional findings in detail, since a few major criticisms will illustrate the authors' general method of dealing with these topics.

The only result for the Merrill-Palmer test which approached statistical significance (critical ratio of 2.5) was a difference of nine IQ points in favor of the preschool group. How much of this is due to differences in rapport is unknown. In general, the analysis of the data on the Merrill-Palmer leads the authors to make strong claims for the effect of preschool, but when, for instance, one finds the statement that "the preschool subjects gained 21.5 points, while the control subjects lost 1.1 points" (p. 67), one wonders why the statement wasn't made to read: the *four* preschool subjects gained 21.5 points, but *one child contributed 11 points to this average*, while *seven* control subjects lost 1.1 points. This is just one of many unqualified statements in this monograph which are apt to mislead the hurried reader.

Despite the fact that the largest critical ratio for the language achievement data was only 1.5, the summary to the section is replete

with positive statements regarding the superiority of the preschool group. Turning to the results for vocabulary, we find that "the preschool group clearly excelled the control group" (p. 97) in vocabulary quotient, but nowhere prior to this statement can one find a single critical ratio, either large or small. Furthermore, the vocabulary quotient used is a highly questionable concept in that it was obtained by dividing the median score made by the orphanage children by the median score made by normal or average Iowa City children of the same age. That the authors did not appreciate the artifacts which result from such a quotient is evident from the following quotation: "It is not clear why the quotients rose with age [from about 20 at age 2 to about 60 at age 5] in the control as well as in the preschool group" (p. 97). A few pages later one finds vocabulary and language achievement quotients being compared and the paradoxical conclusion that "the course of vocabulary development and of language achievement was in opposite directions" (p. 118). Apparently the authors were unaware of the fallacious nature of ratios based upon arbitrary score points. It may be that neither the numerator nor the denominator of such quotients represents anything remotely like a deviation from a real zero point. Make the vocabulary test either easier or more difficult and, presto, the quotients will bounce about.

The chapter on general information contains a statistical "believe it or not": a critical ratio for the difference between means based on N's of 1 and 2; another for N's of 1 and 3; and still another for N's of 2 and 4 (p. 126).

In concluding this discussion of "A study of environmental stimulation" we quote from the authors' final chapter: "Taken all in all, the preschool exerted a profound influence upon the children during the period of preschool enrollment"; "the effect of long residence for the control group was that of tending to bring all children, regardless of initial intelligence classification, to high grade feeble-mindedness or borderline classification"; "the greatest decreases . . . arose for children originally of average intelligence who became feeble-minded." The authors say that these and other statements in the final chapter "may seem unbelievably extreme." We agree. In view of the questions and criticisms which we have raised, we are prepared to make a few statements which may seem unbelievably extreme to the authors and to psychologists and educators who may have uncritically accepted the claims made on the basis of this study. In the first place, there is not a single finding in regard to the influence

of preschool upon mental development which could not be explained on the basis of rapport. In the second place, a critical study of the statistical jugglery reveals that differences in rapport need only be invoked to explain slight, statistically insignificant findings. And finally, the authors are guilty of continually playing up unreliable differences and ignoring not only alternative explanations, but also those parts of their data which do not fit with the environmental hypothesis.

CHILDREN IN FOSTER HOMES

The major part of Skodak's monograph (12) is devoted to a study of 154 foster children who had been placed prior to the age of six months. Of this group, 140 were illegitimates. After at least a year's residence in the foster homes all were tested on either the Kuhlmann or the Stanford Revision of the Binet, and some two years later retests were given. The Kuhlmann was used with children of ages $3\frac{1}{2}$ or younger, while the Stanford was employed with the older children. The median age at first test was one year, seven months, and at the second test four years, one month. The average IQ of the true mothers was 87.7, their average education, 10 grades completed; the average education of the true fathers was also 10 grades completed, and their mean occupational status was 5.4, or .6 of a class below that for the general population. The foster fathers and mothers had, on the average, finished the twelfth grade, and the foster fathers rated 3.1, or 1.7 points better than the general population in regard to occupational status. Let us first consider the major finding of the study, namely, that 154 children, the offspring of parents assertedly much below average and who were placed in superior foster homes prior to the age of six months, were found after a year or more of residence to be above average in intelligence (mean IQ, 116 on first test; 111.5 on second test). This has been hailed by Skeels (8), who also reports on the same study, as "unexpected," and Stoddard speaks of this finding as a "shock to our expectations" (18, p. 47).

One might have anticipated that such shocking results would have led to a close scrutiny of the data rather than hasty acceptance. Just what are the facts?

Let us first examine the claim that "on the basis of information on the intelligence, occupation, and education of the mothers, and the general social status of the true families, it may be stated that on the whole the true family background of these children was inferior to that of the general population on these criteria, which are usually

considered to be indications of ability and intelligence" (p. 102). The true parents are characterized by Stoddard (18, p. 48) as "poor stock." In contrast, the foster homes are described as superior to the average. Now to the evidence.

On the basis of Stanford-Binet tests of only 80 of the 154 true mothers, an average IQ of 87.7 was found. A chronological age divisor of 16 was used; had 15 been used, the average would have been 93.5. The mothers were tested by "various individuals" at three institutions, and "so far as could be determined" the Stanford-Binet was used "either in the completed or abbreviated form." Nothing is said about the expertness (or inexpertness) of the examiners, nor are we told when the tests were given. If, as presumably was the case, the tests were given just before or just after the birth of the illegitimate child, the results would be highly questionable. One would need to have a sublime faith in numerical test scores to take much stock in IQ's determined at a time of such profound emotional stress. The mean education of 144 of the true mothers was 9.9 grades completed, a value which is said to be below the average for the general population.

The intellectual level of the true fathers is inferred from the educational level attained by 88 of the 154 fathers and from the occupational status of 110 of the 154. The mean occupational status was 5.4, or .6 of a point (one-fourth of a sigma) below the general population mean of 4.8. It can thus be seen that the true fathers for whom information was available differed but little from the generality with regard to occupational status. When one considers that many of the fathers were doubtless young (nine were still students, hence unclassified) and therefore had not reached their ultimate occupational level, and when one further considers the likelihood that the 35 unknown fathers were above average, it seems doubtful that the occupational level of the true fathers was inferior. As to their educational achievement, it is reported that the mean grade finished by 88 known fathers was 10.2. This and the corresponding figure for the true mothers form part of the basis for the claim that the parents were subaverage.

At this point it might be well to do what Dr. Skodak did not do, i.e. find data on the average education of the generality of adults. In a bulletin of the U. S. Department of Education it is stated that "the median education in 1934 is only completion of elementary school" (32, p. 14), elementary school being defined as up to and including the eighth grade. This figure is for all U. S. adults and is likely lower than that for adults of ages corresponding to the true

parents. No information is given, however, as to the ages of the parents, but it seems safe to guess that their ages ranged from about 16 up, with a majority under 30. The national school survival figures of Foster (4) indicate that those who were in the fifth grade in 1924 completed about nine grades. This figure agrees with that reported by Bell (1) for the median grade completed by 10,898 youths, 16 to 24 years of age in 1930. This sample was carefully chosen to be representative of the State of Maryland. Even allowing for the fact that the above findings may not hold for Iowa, we are compelled, on the basis of these three sources, to question the assertion of Skodak that the education of the known true parents was below the general adult population level. Perhaps they were actually superior.

Aside from the fact that the mean grade finished by the known fathers is above that for the generality, we have to inquire about the educational achievement of the 66 fathers for whom this was unknown. The best that one can do here is to make a conjecture, then check this with the opinion of others. Our guess is that the "unknown" fathers of illegitimate children are apt to be intellectually superior to known fathers, the intellectual superiority being a factor in their remaining unknown. The reader of this paper can judge for himself the reasonableness of this conjecture.

The above considerations lead the writer to believe that the true fathers were probably above average and that the true mothers were at least average. That the foster homes were above average cannot be doubted, but it has not been demonstrated that the true parents were so far below average as to provide a genuine set for a shock to one's expectations regarding the IQ's of the children. Furthermore, the shock might have been somewhat lessened by reference to norms for the Kuhlmann-Binet (85% of the first tests were Kuhlmann's). According to Kuhlmann (6, p. 13), children of 6, 12, and 18 months of age average 115, whereas, according to Skeels (8, p. 37), the foster children examined prior to 24 months averaged 119. For ages 2 and 3, Kuhlmann gives an average of 107, while Goodenough (5, p. 40) finds an average of 105. The average for the two-year-old foster children was 108. It is therefore possible that these foster children did not really score higher than the average.

Thus, when we consider the intellectual background of the known true parents, the possible level of the unknown parents, and the failure of the children to exceed appreciably the averages found for more nearly unselected children, we are forced to conclude that the

intellectual level of the children is not above that to be expected from their parentage.

The above discussion has been concerned with the major finding based on a foster group placed prior to six months of age. Questions of more or less importance can be raised concerning several points in connection with the treatment of the data for this foster group. There is an indiscriminate mixing of Kuhlmann and Stanford-Binet test results with no thought to the generally admitted faults of the Stanford Revision at the preschool levels. Throughout the discussion, the characterization of the true parents is made to sound as if information were available on all. The difference between the means, 116 and 111.5, for the first and second tests on the children is said to be "slight" (p. 56), although the difference, when the complete formula for the standard error of the difference is used, happens to be 4.5 times its standard error instead of the reported value of 2.99, which Skodak obtained by ignoring the correlational term in the standard error of the difference formula. No concern is shown for the unreliability of testing at the tender age of one year, nor does the author ever seem to appreciate that what is called "intelligence" at these extremely young ages may be decidedly different from what is measured by the Stanford-Binet at ages 4 or 5. She contrasts her findings with researches on older children as though the measures of intelligence for the several ages had been proven comparable.

The correlation of mid-true-parent education with the first test IQ was .08 and with the second test IQ, .33. It is said (p. 78) that at the time of the second examination the correlation had increased "slightly," but was still "substantially" below correlations reported for own-parent-child comparisons. At this juncture she might have reproduced the correlations for own children, given earlier (p. 67), as reported by Goodenough, .35; by Burks, .27; and by Leahy, .48 and .50. Apparently the r of .33 is not "substantially" below that for own-parent-education child IQ relationships. Perhaps the hereditarian could here rescue something from Skodak's study which might tend to support his hypothesis.

On page 84 it is reported that the correlation between the IQ of the mothers and the IQ of the children on the first test was .06, and on the second test, .24. To show that these low figures are consistent with values previously reported in the literature, Snygg and Skeels are cited, but no hint is given that Skeels's data were based on 147 of Skodak's 154 cases. No mention is made of the fact that

the r of .24 was attenuated (1) by unreliability of tests at young ages, (2) by the mixture of Kuhlmann and Stanford-Binet tests, and (3) by the circumstances under which the mothers were tested. Parenthetically, it might be noted here that Snygg (13) ignores a selective factor: the mothers who had passed the high school entrance tests were not included. This definitely restricted the range, and consequently reduced the parent-child correlation by an unknown amount.

It was also found that the older children in foster homes of superior occupational status had IQ's some 12 points higher than those in lower homes. This finding and the correlation of .18 between mid-foster-parent education and second test IQ are of reduced significance in view of the admitted selective placement as shown by the correlation of .30 for the education of true with foster parent, and as evidenced by Skodak's Table 9 (p. 74). From this table we note that there is a correspondence between foster-father's occupational classification and the following: true-father's occupation, true-mother's education, and IQ of true mother. It can be determined from the data given that the correlation ratio for true-mother's IQ on foster-father's occupational status is .35. This definitely indicates selective placement.

Thus, when the obvious selective placement is taken into account, it cannot be claimed that the child's IQ is causally related either to the foster-parental education or to occupational status. This factor of selective placement seems to have been entirely forgotten during most of the discussion and especially when points 2, 3, 5, 6, and 8 in the final summary (pp. 128-129) were written.

A highly questionable finding is the correlation of .49 between final IQ's of the children and an inventory which was devised to provide a measure of the intellectually stimulating value of the environment furnished by the foster homes. Obviously, a part of the obtained correlation can be accounted for on the basis of selective placement. Furthermore, examination of the items in the scale reveals that at least seven of 22 items pertaining to the nonphysical aspect of the home environment may actually be a reflection or function of the child's IQ rather than a producer thereof. For instance, does the child who spends a considerable amount of time reading thus raise his IQ, or does this activity merely reflect a high degree of intelligence? Does conversation of a child with adults lead to increased intelligence or result from native intelligence? That the factors measured by this inventory are not instrumental in changing intelligence is evidenced

by the finding that those who gained 6 points or more from the first to second test were in homes with a mean inventory value of 85.2 as compared to 85.0 for the homes of those who lost 6 to 15 points and 81.7 for the homes of those who lost 16 or more points. (The S.D. for the scale was 14.3.)

From Skodak's study of the IQ's of 16 children of claimed feeble-minded mothers flow the following momentous conclusions: "Thus mother's intelligence appears to have little if any relationship to or influence on the mental development of a child who is removed from her care in early infancy" (p. 91), and "The mental development of children of feeble-minded mothers and the most inferior true-family backgrounds is indistinguishable from that of children whose mothers are not feeble-minded" (p. 104). Wishful thinkers will accept such statements on faith, but others will insist on additional information concerning the intelligence and testing of the mothers and concerning the "other criteria" by which the mothers were judged feeble-minded. Certainly, only the careless will accept the statement that the mean education of the true fathers was ninth grade, but not even the critical will glean from this monograph the fact that information on education was available for only seven of the 16 fathers of these children. This important bit of information was found in another report. More data on the tests and the testing of the children would also be needed for a proper evaluation, and the nongullible prospective foster parent might also ask for data on more than 16 cases before accepting Professor Stoddard's dictum, based on these 16 cases, that "the children of definitely moronic mothers and laboring class fathers, if placed early in good foster homes, will turn out to be above average in mental ability" (18, p. 49).

In closing these comments on the foster children study, we refrain from doing the obviously needed thing—recasting Dr. Skodak's many unsubstantiated conclusions regarding the potency of environment. It is clear that her findings not only fail to support the environmental hypothesis but that they are, in fact, entirely consistent with the hereditarian viewpoint.

OTHER STUDIES

The first two papers of Wellman (21, 22) set the pace for those to follow. Here we find the beginning of the indiscriminate mixing of Kuhlmann and Stanford-Binet IQ's; the first of many analyses in which changes are related to initial IQ level; the abundant use of

graphical presentation and failure to give essential tabular material; and gross misuse of percentiles—IQ's are converted into percentiles, then the percentiles and percentile gains are averaged. How much distortion has been introduced *via* percentiles will never be known unless the original data are reworked.

In Wellman's third paper (23), data are presented to show that IQ changes are not only related to preschool attendance (as claimed in the first two papers) but also to the type of school. The evidence in the first two parts of this third study is in line with claims made in the two earlier studies to the effect that IQ's tend to increase during preschool attendance. Aside from the points raised in the last paragraph and the additional fact that the greatest gain takes place from the first to the second test, it is difficult to criticize this finding on the basis of the data given. Perhaps it may sound unfair to make the general statement that the presentation of the data in these early papers is such as to annoy the reader who is anxious to evaluate critically the results. This is particularly unfortunate in that those who would prefer to accept the claims, but who are cautious, may find the treatment of the data and the control of pertinent variables somewhat obscure.

In the third part of this third study, information is given, then ignored, to the effect that a group of non-preschool children "had had several infant examinations" prior to their first Binet. One wonders, in the absence of information, to what extent experience with infant examinations was an additional uncontrolled variable in these early studies. In so far as the results for this particular non-preschool group are concerned, we have here a possible explanation for their failure to gain from their first to second Binet test. The fact that their initial mean IQ was higher than that for the preschool group is interpreted as indicating that they were rather highly selected, but why stop at this when it might be possible to check on the selection? Surely, information was available on the education and economic levels of the parents. We suspect, however, that their initial test scores were high because of the rapport built up by frequent prior infant examinations. For this group, passing from the first to second test was comparable to going from, perhaps, the fourth to the fifth test for the preschool group, who, it will be recalled, made their greatest gains from first to second or third tests.

Wellman's fourth paper (24) cannot help being a nightmare to statisticians who have for so long held the position that averages and correlations involving percentiles are fallacious. If Dr. Wellman has

analytic proof to the contrary, the statistical world has a right to examine that proof. There are a few other points in this paper that seem questionable. In Table 2 we learn that the correlation between initial IQ and years attendance is .04, but, when it is found in Table 4 that the long-attendance group had an initial mean IQ 6.8 points higher than the short-attendance group, a difference which yields a biserial r of about .27, error can be suspected. In Table 4 one also finds the omission of two disturbingly large critical ratios—the corresponding lower values *were* included in Table 3. These omissions have to do with differences in initial IQ's for the long- and short-attendance groups. These differences are large enough to make one question the reality of the later differences on American Council on Education test percentiles. In fact, when the groups are equated on the basis of initial IQ's (Table 5), the critical ratios drop to 1.9 or less. This mere fact does not lead the author to qualify the summary: "Long attendance children (six or more years in the University schools) consistently made significantly higher scores than short attendance children (one to five years) of equal initial ability" (24, p. 136).

It is in this paper that we note the beginning of the stunt of showing curves for *selected* individuals, and it is here that we find the first mention of the now much-publicized cases who gained from average to the genius level (because of the University school system?). Here one finds certain maximum, and startling, changes pointed out. The greatest change is an increase from 98 at age $3\frac{1}{2}$ to 167 at age 5. This gain is stressed, but the reader is not reminded of the pertinent fact that this individual dropped to 143 at age 10 or to an average of 148 for four tests from ages 9 to 12. Another jump from 89 to 149 is pointed out, but the subsequent drop to 130 is not specifically mentioned. These large individual gains from initial Stanford-Binet tests at age 3 and $3\frac{1}{2}$ may have some meaning, but when it is noted that big gains of 30, 69, and 28 points occur prior to age 6, one becomes somewhat skeptical. In view of the fact that individual cases were found among the data of the orphanage preschool project which showed changes opposite to the cases selected by the authors, and for other reasons, the present writer feels sure that cases among the University school group could be found which would show decrease. In a later paper we find Wellman saying, with regard to four children who showed marked gains, that "these children were not atypical but are representative of a fairly large group" (25, p. 98). Until supporting evidence is given—and it

cannot be found in any of her papers—we are compelled to characterize this statement as a gross exaggeration.

We turn next to the recent monograph by Wellman on "The intelligence of preschool children as measured by the Merrill-Palmer scale of performance tests" (26), but our remarks will be confined to the chief conclusion of that part of the study which has to do with the effects of preschool attendance on Merrill-Palmer scores. The main conclusion is stated as follows: "From these various analyses a sufficient number of positive and significant differences was obtained to justify the conclusion that preschool attendance materially affected ability on the Merrill-Palmer test" (p. 77). It is said that the IQ method of scoring reflected these changes more clearly than sigma scores or percentiles, so let us examine the IQ evidence for the above conclusion. It should be noted that different methods of scoring the test results for the same children will not add to the sampling significance of the results. Accordingly, if we demolish the finding with respect to IQ scoring, the method which "reflected these changes more clearly," it should not be necessary to state here the detailed argument which could be produced to explain away the findings with regard to other scoring schemes.

Let us now evaluate the data upon which the above-quoted conclusion is based. It is reported (Table 11, p. 41) that 72 "cases" gained 9.1 IQ points from fall to spring tests, the gain yielding a critical ratio of 3.37 (one of two significant ratios for results dealing with effects of preschool upon Merrill-Palmer IQ), and that 46 "cases" gained 3.7 points from spring to fall tests, a gain which is 1.09 times its standard error. Now there are three important questions, or issues, to be raised here. First, in determining the significance of the gains, the correlation term in the standard error of the difference formula was ignored. This is a frequent error in these Iowa studies, and one which does not always lead to a conservative statement of significance, *e.g.* see Skodak's (12, p. 56) "slight" change which, properly evaluated, is 4.5 times its standard error. The gains of 9.1 and 3.7 IQ points given above are, on the basis of the given N's and r's, actually 5.32 and 2.10 times their standard errors, respectively, for fall to spring and spring to fall. Second, the 72 "cases" given are not 72 different children, nor can one be sure that the 46 "cases" are 46 different children (see the bouncing N's in Table 10, and subsequent discussion on page 40), and therefore we again have another instance in which one of the assumptions of sampling has been violated. Since the true N's are not given, we

cannot make an adjustment for this incorrect use of error formulas; we can only state that such a correction would tend to increase the standard errors and hence lower the critical ratios. Third, Dr. Wellman has been content to stop with evidence for a significant gain for one group and an insignificant gain for the other, but in reality an adequate statistical treatment must evaluate the *difference* between the gains.

Using the error formula proper for correlated means, but with no adjustment for the real (and unstated) N's, we have the gain from fall to spring as 9.1 ± 1.71 and the gain from spring to fall as 3.7 ± 1.76 ; then the difference between the gains, 5.4, is readily found to be 2.2 times the standard error of the difference. This ratio is too high because of the use of inflated N's. In order to make a justifiable comparison so far as the N's are concerned, we note the results (Table 11) for 42 children having a first test in the fall and whose gain to spring of 10.7 IQ points is reported to be 3.35 times its standard error (the second of two significant ratios for results dealing with effects of preschool upon Merrill-Palmer IQ), while 20 children having a first test in the spring made a gain of 5.2 points (.97 times its sigma). In evaluating the significance of the *difference*, we first recompute the standard errors of the gains by the formula which includes the *r* term. The needed *r*'s are not given; presumably, they will not differ much from those reported for the total groups, as given in Table 14 (p. 54). The difference between gains, 10.7 minus 5.2, is found to be only 1.56 times its standard error. The fact that this gain is consistent with that found for the larger groups adds nothing to the significance of the finding, since it is based on subsamplings of the larger groups. Thus, it is seen that the author's conclusion regarding the effect of preschool attendance on Merrill-Palmer IQ's resulted from faulty statistical treatment of the data.

In a paper by Skeels and Fillmore (9) it is concluded that the longer subaverage children remain in their underprivileged homes, the lower are their IQ's. This conclusion was based on a comparison of means for older as opposed to younger children, but the significant drop with age was not properly evaluated in that no allowance was made for the fact, reproduced in their paper, that the 1916 Stanford Revision yielded a negative correlation for IQ with age for unselected children. An adequate statistical analysis must determine the significance of the difference between the drop for their group and that for the unselected group. This can readily be accomplished by either of two methods. One can determine from their data that the mean for

ages 12 to 14 combined is 11.8 points lower than the mean for ages 5 to 7 combined. The corresponding figure for unselected cases is 7.5. The difference, 4.3, between the differences happens to be only 1.6 times its standard error. The second method, which is preferable in that it utilizes all the cases from 5 to 14, inclusive, is to compare the slopes of the two regression lines for IQ on age. From the data given one cannot ascertain exactly the two regression coefficients, but an excellent approximation can be obtained by fitting lines to the age means, weighted according to their respective N's. The difference between the two regression coefficients, *i.e.* the slopes, so computed is 1.8 times its standard error. When we also consider the data in Wellman's (21) earlier study, which showed that above-average children in above-average homes drop about 10 or 11 IQ points with age (ages 5 to 7 combined compared to ages 12 to 14 combined), it appears that the conclusion of Skeels and Fillmore is entirely unwarranted.

IQ CHANGES AND REGRESSION

The analysis of changes in IQ according to initial IQ level has been persistently pursued in all the Iowa studies so far mentioned, and has been given such prominence in the monograph by Crissey (3) as to lead Professor Stoddard in his foreword thereto to point out the main finding as being the fact that "changes in brightness tend to be related to the general IQ level of the group: the relatively dull move upward and the relatively bright show losses." In Wellman's study of the Merrill-Palmer scale we find that IQ's on this scale are analyzed in terms of Binet IQ level, and vice versa, with the resultant finding "that children who receive low Merrill-Palmer scores can be expected to receive higher Binet scores than their Merrill-Palmers, and children who receive low Binet scores can be expected to receive higher Merrill-Palmer scores . . ." (26, pp. 99-100). This is just the result which any competent statistician would have confidently predicted, providing he had been told the fact that the standard deviations for Binet and Merrill-Palmer IQ distributions were approximately the same and that the correlation between the two sets of scores was low (say, less than .80).

Our main concern here, however, is with the test-retest changes according to initial IQ level. Except for a couple of instances, all of the Iowa analyses on this point have yielded results consistent with the finding, cited above, of Crissey. These findings can be summarized in correlational terms by saying that there is a negative

correlation between changes expressed as gains and initial IQ level. So stated, one type of divergence from Crissey's result can be brought into line. We refer to the situation where gains are shown at all initial IQ classifications, but are still inversely related to initial level; or to the situation where losses occur at all levels, but those initially lower lose less than those initially above the group average. The other exception to the Crissey result is to be found in Wellman's analysis of retests at a week's interval on the Merrill-Palmer. It was found that "the amount of gain increased with successively higher initial IQ classifications" (26, p. 30). The explanation of this in terms of practice effects, with superior children profiting more, is acceptable to the writer.

That IQ changes from one test to a later retest are related to IQ level on the first test cannot be denied, but when an environmental explanation is advanced, we begin to feel that the environmental hypothesis is being overworked. It is suggested by Dr. Wellman (26, p. 53) that the superior do not find their environments sufficiently challenging and hence tend to lose on retest, and Crissey (3, p. 21) suggests that those initially below average tend to gain because they find their environments stimulating. Dr. Wellman has maintained that ordinary statistical regression has nothing whatever to do with the inverse relationship between gains and initial IQ level, but we are forced to readvance the concept of regression, which has been labeled by Dr. Wellman as an "hypothesis," despite the position of qualified statisticians that regression is a "fact."

Before turning to Dr. Wellman's arguments against regression, let us consider an unpublished finding of the present writer. Fifty-four children of initial IQ's between 140 and 149 lost an average of five points (CR of 3.00) on a retest. Shall we attribute this loss to lack of environmental stimulation? To do so would stretch the imagination of even the most hopeful environmentalist, since this loss occurred within a week. The loss represents nothing more than statistical regression as we pass from Form M to Form L of the New Stanford-Binet, and is due solely to errors of measurement. A similar loss occurs when we pass from Form L to Form M, and gains occur for those classified as inferior on either form and tested a week later on the other form.

It should be explicitly noted that we are not claiming that all the gains for the inferior and all losses for the superior reported by the Iowa investigators are explicable on the basis of errors of measurement. However, the reliabilities of the Kuhlmann and Stanford-

Binet and the Merrill-Palmer at ages 2 to 5 are not high enough to preclude the possibility that a large part of such gains and losses is due to errors of measurement. The remaining portion of the changes, differential with regard to IQ level, from test to retest six or more months later may be attributable to differences in maturation or, conceivably, to differences in environmental stimulation. But before we accept the hypothesis that losses for those initially above the group average are due to a lack of environmental stimulation, and that gains for those below the group average are due to the environment being stimulating for them, we must ask about a further result which should follow if this hypothesis is to be tenable. If the hypothesis were true, we would expect a reduction in variability from initial to later test, but this happens not to be the case, not even for the control group in the orphanage. Perhaps the Iowa people, who are quite adept at finding an environmental explanation for all changes, can produce one to account for the fact that the number of people in the several IQ classifications is approximately the same for later tests as for the initial test, despite the fact of differential changes.

If these investigators should insist on the correctness of their concept concerning the "stimulating value of the group," they must explain the fact that individuals classified above average on the basis of a final test will, in general, have had lower initial IQ's, while those below average on the final test will have been higher on the initial test. Perhaps the "stimulating value of the group" has acted retroactively! But we hasten to point out that this merely describes the fact that some (a large number) of the individuals initially above average do gain, while some below average do lose—gains and losses which occur in spite of the supposed stimulating value of the total group.

Let us look at the problem from the analytical viewpoint. Given an initial IQ, x_1 , and a retest IQ, x_2 , and let the gain be defined as $g=x_2-x_1$, then it can be shown by easy algebra that the correlation between initial IQ and gain is

$$r_{1g} = \frac{r_{12}\sigma_2 - \sigma_1}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2r_{12}\sigma_1\sigma_2}}$$

from which it can readily be seen that, unless the variability increases from first test to the second or later test, the correlation between gains and initial IQ *must* be negative, since in practice r_{12} will never be unity. This, of course, does not explain the negative correlation; it merely indicates that the Iowa investigators have gone to an enor-

nough losses of the st six ration. But above n, and environment result of the from en for , who or all per of ne for changes. their must basis those initial acted scribes tially and f the Given defined corre- cases veen ever ion; nor-

mous amount of work to demonstrate a fact which even a mediocre statistician could prove analytically in less than five minutes. Perhaps we should not object to empirical demonstrations, but when these lead to such fallacious conclusions as that concerning the "leveling effect" on intelligence of residence in an orphanage (discussed earlier in this paper), it is time for the artifactual nature of the finding to be reviewed.

It is difficult to understand how Dr. Wellman could discuss the relationship between Merrill-Palmer and Binet IQ's, or study her graphs (26, p. 98), without recognizing regression, but no mention is made of this fact when she speaks of those with high Binets having lower Merrill-Palmers, etc. Earlier in this same monograph, however, we do find Dr. Wellman discussing and rejecting regression as it affects test-retest changes. Her argument is so pertinent that we reproduce it here in full:

"Another hypothesis is that commonly referred to as regression towards the mean. According to this hypothesis very superior children are expected to lose and below average children to gain, both extremes approaching the mean. This phenomenon when observed has been interpreted at times as purely statistical, the chances of loss of high-scoring children and the chances of gain of low-scoring children being automatically greater. At times the phenomenon has received a biological interpretation of the tendency of the organism to veer toward the general level of the race. There are two difficulties in acceptance of the regression-towards-the-mean hypothesis here: (1) the facts of change on retest at one week do not fit the expected trend, and (2) the winter and summer groups are selected superior groups. Instead of gaining, the children at 0.0 sigma score should not have changed, and the children at 0.5 and 1.0 sigma score should have lost" (26, p. 53).

Following this rejection of the regression "hypothesis," Dr. Wellman proposes the earlier-mentioned explanation of the differential gains as being a matter of environmental stimulation. Let us examine her argument. In the first place, the regression phenomenon is, of course, in this case purely statistical as opposed to any notion of biological regression. As we understand the latter concept, it has to do with progeny as compared with parent, and, as such, the latter concept is absolutely not applicable to the test-retest situation. The first of the two difficulties has already been mentioned earlier by the present writer as an exception to Crissey's finding, and at that time we accepted Wellman's explanation in terms of practice effects—the superior are better able to profit therefrom, and consequently we have a factor of such potency as to overcome the ordinary statistical

regression. This, in terms of the formula given above, really means that σ_2 had to be greater than σ_1 (about $1.2 \sigma_1$).

In regard to the second difficulty, it should first be noted that the changes over a six-months period for the selected superior groups do exhibit regression, as is evidenced by Wellman's Table 13 (26, p. 52) and by the negative correlations of .38 and .43 between changes and initial IQ. This regression, however, is about the means of the groups concerned and not about the population mean. This should not be surprising, since their original classification as "selected superior groups" was not made on the basis of the first test. They earned this classification *via* whatever factors operate to fill preschools with superior children. This differs from, for example, Terman's selection of "gifted" children as those above 140 IQ and the consequent drop (regression) on later tests. Perhaps an analogy may help us understand why Wellman's selected superior children did not, and need not, as a group, regress toward the population mean from test to retest. Suppose we choose a group of eight-year-old boys of Swedish extraction and determine that their mean height was one-half a sigma above the mean for the generality of American boys of that age. After a period of eight years we remeasure them; the correlation between the two sets of measures will not be high—changes in the relative standing of the individuals will have occurred, but we would not expect the group as a whole to be nearer the mean of all 16-year-old American boys than one-half a sigma (this second sigma must, of course, be based on 16-year-olds). We would expect regression *within* the Swedish group, but this would not tend to reduce their superiority nor would it lead to a reduction in the absolute or relative variability of the group. In contrast, if, on the basis of measurement, we had chosen a group of boys as above average, we would find a general tendency for the group, so selected, to be nearer the universe mean upon subsequent measurement.

Perhaps the point can be better illustrated by a mental test situation. Suppose it has been established (1) that the father-child IQ correlation is .50 for each of the child age levels 6 to 14, (2) that the average IQ for each age level and for fathers is 100, and (3) that the S.D.'s for each level and for the fathers is 16. Now let us select for study those six-year-old children whose fathers have IQ's of 80; the average IQ of these children will be 90, the S.D. will be 13.86. It should be obvious that if we retest these children at age 14 they will again average 90 (since the average IQ for all 14-year-old children of fathers with 80 IQ will be 90). That is, the test-retest

regression will not have brought them nearer the population mean, but, within the group, regression about the mean of 90 will have taken place. In other words, an inferior or superior group will not move toward the general mean on a retest unless they have been selected as inferior or superior on the basis of an initial test.

Aside from the general, and easily anticipated, finding of Crissey that changes are inversely related to initial IQ level, it is of some interest to follow through a different type of analysis which purports to substantiate the hypothesis that "the rate of mental development of a child in an institution designed for normal and dull-normal children should vary from that of a child of similar mental ability in an institution designed for the feeble-minded" (3, p. 52). To check this hypothesis the method of matched groups was used, and one of the matching criteria was that the individuals had to be within three points in IQ on initial test. Since the average initial IQ for the individuals in the institution for dull-normals was about 85, as compared to an average of about 62 for those in the institution for feeble-minded, it follows that in order to equate on initial IQ it was necessary in general to match children from the lower end of one distribution with children from the upper end of the second (or feeble-minded) distribution. This type of thing tends definitely to capitalize on errors of measurement, with the result that the individuals drawn from the first group will regress upward, while those drawn from the second group will regress downward on a later test. It is not surprising, therefore, when Crissey (3, p. 53) finds that, for four different sets of matched groups, the groups selected from among dull-normals gain and those selected from among feeble-minded lose. A significant difference in changes does not, of course, preclude the possibility that the changes were due to errors of measurement *via* regression.

There are two methods for making allowance for the regressive effect of measurement errors on IQ changes of the sort being discussed in this section. We can make our classification on the basis of regressed ($x_{\infty} = r_{11}x$) initial IQ's and then determine the changes (gains or losses) from these regressed scores. By this method any mean gain on the part of the initially low or mean loss on the part of the initially high will not be due to errors of measurement. The second scheme for eliminating the effect of errors of measurement is to state the relationship between initial IQ and changes in terms of the correlation coefficient and then correct this for attenuation. It has long been known that this correction will *reduce* (bring nearer

zero) the correlation, or, conversely, that the effect of measurement errors is to produce a negative correlation between initial scores and gain (19).

In closing this rather lengthy discussion of regression we agree with Dr. Wellman that regression "is really more of a descriptive than an explanatory term" (26, p. 32), but it does not follow from this that pages of tedious analysis need to be devoted to pointing out that those initially high tend to lose and those initially low tend to gain. This can readily be inferred from the test-retest correlation and the sigmas and means for a given group. Neither does it follow that regression from test to retest exemplifies a leveling effect. Furthermore, no definite conclusions can be drawn from the many analyses showing losses for those above the group average and gains for those below until due allowance is made for the portion of these changes which is attributable to regression because of errors of measurement.

SUMMARY AND CONCLUSION

So far in this critical examination of the Iowa studies on IQ changes we have been content to raise specific questions concerning definite methodological and statistical inadequacies. In summarizing, we find it necessary to make a few general statements, the validity of which can be judged by the reader. In brief, we have found much of the supposed evidence for environmental influences on the IQ to be entirely nonexistent. We have cited instance after instance, and have left unmentioned many more examples of minor importance, in which the findings have resulted from either uncontrolled factors or erroneous statistical treatment or both. We have noted, but not stressed, the fact that these studies are replete with misleading, *i.e.* not properly qualified, statements regarding the influences of school and environment. We have said little about the fact that insignificant findings in favor of the environmental viewpoint have been constantly played up while contrary findings have been ignored. We have also noted a disturbing tendency to a dramatic use of selected cases, falsely claimed to be typical, and a simultaneous disregard for other cases which would just as dramatically disprove their contentions.

In conclusion, the writer would like to express two personal opinions. First, in view of the fact that we have discovered startling inadequacies in those studies reported in monographic detail and in

view of the fact that it was not until the original data were secured that we were able properly to evaluate—in this case demolish—the evidence based on the orphanage preschool project, we are strongly skeptical as to the dependability of the results which have been reported all too briefly in the shorter papers, by which we mean specifically the earlier papers of Dr. Wellman. Second, if it is the responsibility of the scientist to establish, and the educator to disseminate, truths, then our scientists who have turned educators should take the responsibility for dispelling error, especially that which has been the result of their own hasty promulgation of unverified and largely invalid research results.

BIBLIOGRAPHY

1. BELL, H. M. Youth tell their story. Washington: American Council on Education, 1938.
2. COFFEY, H. S., & WELLMAN, B. L. The role of cultural status in intelligence changes of preschool children. *J. exp. Educ.*, 1936-1937, 5, 191-202.
3. CRISSEY, O. L. Mental development as related to institutional residence and educational achievement. *Univ. Ia Stud. Child Welf.*, 1937, 11, No. 1.
4. FOSTER, E. M. School survival rates. *Sch. Life*, 1936, 22, 13-14; 1938, 23, 265-267.
5. GOODENOUGH, F. L. The Kuhlmann-Binet tests. Minneapolis: Univ. Minnesota Press, 1928.
6. KUHLMANN, F. A handbook of mental tests. Baltimore: Warwick & York, 1922.
7. SKEELS, H. M. Mental development of children in foster homes. *J. genet. Psychol.*, 1936, 49, 91-106.
8. SKEELS, H. M. Mental development of children in foster homes. *J. consult. Psychol.*, 1938, 2, 33-43.
9. SKEELS, H. M., & FILLMORE, E. A. The mental development of children from underprivileged homes. *J. genet. Psychol.*, 1937, 50, 427-439.
10. SKEELS, H. M., UPDEGRAFF, R., WELLMAN, B. L., & WILLIAMS, H. M. A study of environmental stimulation. *Univ. Ia Stud. Child Welf.*, 1938, 15, No. 4.
11. SKODAK, M. The mental development of adopted children whose true mothers are feeble-minded. *Child Devolpm.*, 1938, 9, 303-308.
12. SKODAK, M. Children in foster homes: a study of mental development. *Univ. Ia Stud. Child Welf.*, 1939, 16, No. 1.
13. SNYGG, D. The relation between the intelligence of mothers and of their children living in foster homes. *J. genet. Psychol.*, 1938, 52, 401-406.
14. STODDARD, G. D. What of the nursery school? *Progr. Educ.*, 1937, 14, 440-451.
15. STODDARD, G. D. Our children: their intelligence. (Abstract.) *Proc. nat. Educ. Ass.*, 1938, 76, 61-62.
16. STODDARD, G. D. Child development—a new approach to education. *Sch. & Soc.*, 1939, 49, 33-38.

17. STODDARD, G. D. Education for self-realization. *Nat. Parent-Teach.*, 1939, 33, 5-8.
18. STODDARD, G. D. The IQ: its ups and downs. *Educ. Rec. Suppl.*, January, 1939, 44-57.
19. THOMSON, G. H. An alternate formula for true correlation of initial values with gains. *J. exp. Psychol.*, 1925, 8, 323-324.
20. UPDEGRAFF, R. The determination of a reliable intelligence quotient for the young child. *J. genet. Psychol.*, 1932, 41, 152-166.
21. WELLMAN, B. L. Some new bases for interpretation of the IQ. *J. genet. Psychol.*, 1932, 41, 116-126.
22. WELLMAN, B. L. The effect of pre-school attendance upon the IQ. *J. exp. Educ.*, 1932-1933, 1, 48-69.
23. WELLMAN, B. L. Growth in intelligence under differing school environments. *J. exp. Educ.*, 1934-1935, 3, 59-83.
24. WELLMAN, B. L. Mental growth from preschool to college. *J. exp. Educ.*, 1937-1938, 6, 127-138.
25. WELLMAN, B. L. Our changing concepts of intelligence. *J. consult. Psychol.*, 1938, 2, 97-107.
26. WELLMAN, B. L. The intelligence of preschool children as measured by the Merrill-Palmer scale of performance tests. *Univ. Ia Stud. Child Welf.*, 1938, 15, No. 3.
27. WELLMAN, B. L. Guiding mental development. *Childhood Educ.*, 1938, 15, 108-112.
28. WELLMAN, B. L. New tests attack theory of fixed IQ. *New York Times*, July 17, 1938, Section II, 4.
29. WELLMAN, B. L. How the child's mind grows. *Nat. Parent-Teach.*, 1939, 33, 17-18.
30. WELLMAN, B. L. The changing concept of the I. Q. *J. Home Econ.*, 1939, 31, 77-80.
31. WELLMAN, B. L., & STODDARD, G. D. The IQ: a problem in social construction. *Social Front.*, 1939, 5, 151-152.
32. ——. Biennial survey of education, 1933-1934. *U. S. Off. Educ. Bull.*, 1935, No. 2.

939,
ary,
initial
for
enet.
exp.
ron-
duc.,
sult.
d by
Child
1938,
imes,
1939,
con-
Bull.,

REVIEW OF McNEMAR'S CRITICAL EXAMINATION OF IOWA STUDIES

BY BETH L. WELLMAN, HAROLD M. SKEELS,
AND MARIE SKODAK

University of Iowa

The reader may wonder if it is McNemar's scientific self that speaks in his long critique of the Iowa studies of intelligence. Calm, impartial weighing of evidence for and against a conclusion signifies the scientific approach. But McNemar has embarked upon the mission of "demolishing," "destroying," and "exploding" the conclusions of the Iowa investigators.

In analyzing data, investigators are often confronted with alternative methods of treatment. In several instances our choice of method has not appealed to McNemar. However, we have carried out certain of his suggestions, only to find that the "approved" statistical treatment did not change the major conclusions.

Since most of the criticisms made by McNemar can be evaluated only by a detailed comparison with the original monographs and articles, and since this is a time-consuming task, it seems desirable to present here a fairly complete reply. The answer to some of the criticisms is contained in the portion of the studies to which McNemar does not allude.

PUBLICITY

The point of view that the intelligence quotient remains constant has received publicity greater in volume and diversity than that thus far accorded the Iowa studies. Theories of intelligence, whether they point to a static condition untouchable by environmental impacts or to plasticity, are of practical importance to educators, social workers, and parents—to all who are interested in child development. Workers in other fields have a right to be informed of new findings and conclusions that may affect their practices. To withhold such information is to betray a professional trust.

But the "gospel" is certainly not original with Iowa investigators: it was preached by Binet in 1909 (1).

We agree with McNemar that findings should be supported by the best of evidence. We believe, also, that criticism should be directed to the basic issues, methods, and conclusions, and not to minor details. For example, the gain of preschool children while in preschool and the failure to gain while not in school (based on several hundred cases) should not be overlooked. Moreover, a critical re-examination of data on which the concept of a "constant IQ" was based would be illuminating.

THE ORPHANAGE PRESCHOOL PROJECT

We note McNemar's statement that "it was not possible to control more than two or three at a time of the following variables" in the orphanage preschool project (10). We do not claim that all possible factors were perfectly controlled. We do feel, however, that there was reasonable control over the variables mentioned, in addition to the fact that the lives of these children were circumscribed by residence in an orphanage.

Variables 1 (age), 2 (initial intelligence), and 3 (orphanage residence) figured in the original matchings of preschool and non-preschool groups and were thus controlled. Further, the results were presented in terms of changes by 10-point IQ levels. We assume that by variable 3 McNemar means residence in the orphanage prior to the project. If, however, he means residence after the project was started, his statement does not apply, since analyses were presented in terms of length of residence. Variable 4 (actual number of days of school attendance) was always zero for the control group, and thus this variable was controlled. Variable 5 (days of residence between tests or retest intervals) formed the basis of division into subgroups for analysis. Variables 6 (various examiners), 7 (practice effects), and 8 (Kuhlmann-Binet or Stanford-Binet) operated similarly in preschool and control groups. The testing program was identical, the examiners were the same, and the basis for choice between Kuhlmann and Stanford Revisions was the same in both groups. Variable 9 (possible unintentional coaching in preschool) is a poser. It is difficult to see how one can prove that he was innocent of "unintentional coaching" in any situation. Variable 10 (differences in rapport in testing) is to be discussed later.

In presenting his Table I, McNemar states that "no allowance has been made here for certain variables mentioned above." Presumably, he means that the authors were the guilty ones. However,

McNemar is the one who did not make the allowance for the variables. His Table I is a condensation of our Table 2 (10, p. 42), which is in turn derived from our Table 1 (10, p. 41). Our Table 1 gives details on initial IQ, days residence, and days of preschool attendance, which he has apparently overlooked. The other variables were controlled by the original matchings and by the testing program, as stated above.

In order to clear up the point of the effect in this study of using cases rather than individuals, we present here some results of analyses made in terms of individuals. The 11 preschool children with an initial IQ of 80 or above who were in residence the longest (400 days or more) and the 11 control children of similar initial IQ and residence showed the following changes:

	Preschool	Control
Mean initial IQ	92.6	93.9
Mean final IQ	90.4	78.3
Change	-2.2	-15.6

The 11 control children showed a mean loss of 15.6 points. This difference is statistically significant by Fisher's *t* formula. The preschool group lost 2.2 points. This change is not statistically significant. The net difference between these two is 13.4, and this is significant (*P* is less than .01).

Thus, the statement quoted by McNemar can be amended to read as follows: "In two years' time a group of 11 children who averaged 94 in IQ dropped 16 points." McNemar's statement "that such changes as have taken place [are] of doubtful statistical significance" is not correct for the children initially above 80 IQ.

Still a
play

The 10 preschool children and the 11 control children whose initial IQ's were below 80 and who were in residence 400 days or more made the following changes:

	Preschool	Control
Mean initial IQ	72.4	68.9
Mean final IQ	80.6	73.1
Change	+8.2	+4.2

While the gains were not statistically significant for either the preschool or the control group, the trend is consistent.

We stated in the monograph that there was little difference between preschool and control groups in the shorter period of residence, but that the divergence between the two groups became accentuated with longer residence. We present here the successive

changes for all individual children who received a test within each of the three residence groupings. (When a child received two tests within a given residence grouping, the one associated with the longer residence was used.) The changes for the children initially below 80 IQ were:

	Preschool (N=6)	Control (N=7)	Difference
Initial IQ	72.2	71.9	0.3
IQ at 1-199 days	74.7	71.1	3.6
IQ at 200-399 days	79.8	74.4	5.4
IQ at 400 days or more	84.2	77.4	6.8

The changes for the children initially 80 or above in IQ were:

	Preschool (N=7)	Control (N=5)	Difference
Initial IQ	94.9	90.8	4.1
IQ at 1-199 days	89.4	87.6	1.8
IQ at 200-399 days	90.4	81.6	8.8
IQ at 400 days or more	94.1	74.2	19.9

McNemar states that rapport was "one of the most important of the uncontrolled variables." Our examiners were all experienced in giving intelligence tests to children of preschool ages; not one reported difficulty in securing rapport. Rather, the children liked to be tested. This is the general experience with young orphanage children. They crave individual attention. McNemar is confusing a description of the children's behavior in a group situation, as viewed by the teacher, with behavior in an individual test situation. The authors did not "overlook" the matter of rapport, but took great pains to establish satisfactory relations with every child tested.

McNemar further states that "this factor *alone* (italics ours) might serve as an adequate explanation" of divergences. If so, why did the control children who were initially below 80 IQ improve (they gained 4.2 points) while those initially above 80 IQ decreased to the extent of 15.6 points? Could rapport work in opposite directions in the two portions of the control group? Also, if the preschool children became more coöperative, why did rapport not improve for those initially above 80 IQ (they lost 2.2 points)?

All children, preschool and control, were residents of the orphanage and were examined in a familiar setting. We believe that the examiners established a friendly relationship with each child such that responses to test items represented the child's best efforts.

We assume that regression effects are equal in two similar groups. Our presentation in the monograph (10, p. 41) of IQ changes by

10-point IQ levels shows that the changes for preschool and control groups were widely different. The changes reported in the monograph for the long-residence group (400 days or more) were:

Initial IQ	Preschool		Control	
	Cases	Changes in IQ	Cases	Changes in IQ
100 and above	2	-4.5	2	-28.5
90 to 99	8	0.9	10	-17.6
80 to 89	7	1.6	14	-13.5
70 to 79	16	4.8	22	1.0
60 to 69	7	14.2	13	6.5
50 to 59			4	3.5

Reanalyzed in terms of individuals instead of cases, the changes were:

Initial IQ	Preschool		Control	
	Children	Changes in IQ	Children	Changes in IQ
100 and above	2	-4.5	2	-28.5
90 to 99	5	-1.0	5	-12.6
80 to 89	4	-2.5	4	-13.0
70 to 79	7	4.4	7	3.3
60 to 69	3	17.0	3	7.0
50 to 59			1	2.0

The quotation on leveling effect of long residence for the control group, which McNemar found "amazing," referred to final means for groups divided according to initial IQ. The reader is referred to Figure 5 (10, p. 40), entitled *Trend of Changes in Binet IQ With Residence: By 10-Point Classifications*. Every control subgroup with long residence, regardless of whether the initial IQ's represented the 60's, 70's, 80's, 90's, 100's or above, received a final mean IQ between 70 and 79. We referred to this as the leveling effect. (When the data were reanalyzed, using each individual only once, the range of final means was extended slightly, the range then becoming 68 to 83.)

Preschool children of long residence showed a different trend. Only the two lower classifications of preschool children (that is, the groups with initial IQ's in the 60's and 70's) had final means between 70 and 79. The other three groups (that is, those with initial IQ's in the 80's, 90's, and 100's or above) had final means in the 80's, 90's, and 100's. (When the data were reanalyzed, using each individual only once, one final mean fell in the 70's, two in the 80's, one in the 90's, and one in the 100's.)

It is possible for a series of means to approach each other (leveling effect) while the S.D. for the combined groups remains approximately the same. Our "statistical laboratory" cannot claim the

falling off

credit for producing final IQ's in the 50's and 60's. More than one-third of the long-residence control children had final IQ's within these limits, while only one preschool child had a final IQ below 70.

That correlations between the number of days of preschool attendance and changes in IQ are not too revealing is indicated in the following statement:

"It appeared that not only actual days of attendance but the distribution of those days throughout days of nonattendance was important in determining amount of change. Anything less than 80 per cent attendance was not instrumental in producing gain, even though the actual attendance was 150 to 175 days. Conversely, anything below 70 per cent attendance was likely to produce loss, and if 60 per cent or less, the loss was pronounced" (10, p. 56).

There were three preschool children and one control child in the long-residence group (400 days or more) who gained 15 or more points. The greatest losses in the preschool group were 15, 14, and 13 points, respectively. The greatest losses in the control group were 43, 37, 23, 15, 14, and 14 points. Since it was stated that the preschool and control children were not differentiated by short periods of residence, there seems to be no point in McNemar's selection of cases differing slightly from each other.

Losses were no more characteristic of children who were tested at a very young age (prior to 20 months) than of children tested at a later age. For the long-residence children, new calculations have been made, omitting test scores obtained below 20 months. For each child we substituted, as his initial test, the next succeeding test score. The 11 children for whom we previously reported a drop of 15.6 points were now shown to have a mean decrease of 12.2 points. Both decreases are statistically significant. Also, the difference between the change of the preschool and control children remains statistically significant.

We should be glad to have data from any source in the research literature as to the comparability of measures of intelligence at any two ages for which the tests have been standardized.

Six children from the control group were transferred to an institution for the feeble-minded. Of these, three had been in the orphanage at the start of the project. Their initial IQ's were 103, 98, and 73, and their IQ's at the time of transfer were 60, 61, and 62. Since the other three children who were transferred were not in the situation at the beginning, we agree with McNemar that it is better to

rule them out. From the preschool group two children were transferred, neither child having attended as much as 80 days. One child was five years, six months, of age at the start of the project; the other had attended only 56 days, having entered the project only 12 weeks before it ended.

For the "long-residence" children the average initial IQ for the preschool group was 82.9; and for the control group, 81.4. For the complete group the initial IQ averaged 86.6 in the preschool situation and 82.1 in the control.

We are unable to find any reasonable explanation as to why McNemar wishes us to stress the 11-point gain in Merrill-Palmer IQ made by one preschool child. Reference to Table 13 of the monograph (10, p. 67) shows that the changes for the four preschool subjects were, respectively, 43, 9, 11, and 23 points (average, 21.5 points).

McNemar states that the section on language achievement "is replete with positive statements regarding the superiority of the preschool group." Let us quote a few introductory sentences from three paragraphs of the summary in the monograph (10, p. 85): "Increasing losses in language quotient for both groups were found for the three intervals studied, six, twelve, and eighteen months. . . . The results on changes with amount of preschool attendance were unclear. . . . With respect to total amount of preschool attendance up to the end of the periods for which changes were calculated, the results were again unclear."

The results above are based on the Little-Williams Language Achievement Scale, which covers speech sounds, intelligibility, and sentence organization. The following quotation from the summary on vocabulary (10) may be of interest:

"There were distinct differences between preschool and control children in gains on retests. At the end of a five months' period the difference was very slight, but at the end of a period of one year there was practical certainty that the preschool children were higher in vocabulary than the control children. The gain of the preschool children had been nearly double the gain of the control children" (p. 113).

The critical ratio of the difference between the final scores was 4.0.

CHILDREN IN FOSTER HOMES

McNemar's criticisms of the data on the true parents in the Skodak (12) and Skeels (8) studies of children in foster homes may arise, in part, from his lack of familiarity with the problems of

illegitimacy and child placement, the types of true homes foster children come from, and the case records available. Persons who have had first-hand contacts with unmarried parents whose children are placed through public agencies would not classify these parents as average or above along intellectual and social lines. The records on true parents are summarized in the Skodak monograph (12, pp. 38-44).

In view of the absence of measures of genetic constitution (that is, "good" or "bad" stock), the practical measures are limited to intelligence tests, reports of school progress, occupation, and family status in the community.

With regard to paternal history, it should be stated at the outset that there is no way in which a man can be proved the indisputable father of a particular child. In securing paternal history the first, and often only, information comes from the mother. The father may be genuinely unknown, as in the case of girls who have been promiscuous. The pregnancy may have occurred following assault by a stranger. The putative father may have been a chance "pickup"; he may have given a false name. Compared to the mother, he may have been superior, equal, or inferior. The father may remain unknown, not because he was superior to the mother, as McNemar implies, but because the mother was unable or unwilling to give the necessary information. The paternity of the child can be assumed to be known when the information of the mother can be checked and when the alleged father admits the possibility. The fathers in this study that can be placed in this category came from the same socio-economic levels as the mothers.

Information on the maternal histories comes primarily from the mothers themselves. In some cases the data were known and checked. In most cases the data as given by the mother were accepted. It may be that the mothers tended to exaggerate with respect to education and community adjustment.

This study grew out of a clinical program. The data as recorded were the only materials available. Of the 154 children, 114 had been placed by a public agency that was obliged to accept all children, regardless of family background, so far as the capacity of the institution permitted. Thirty-five children had been placed by a private agency that was slightly selective in the families from which it accepted children.

Although illegitimate childbearing is not confined to any one social level, it is probably less frequent among girls from superior homes. If a pregnancy is discovered, the girl in the "better family"

receives more protection; funds may be available to carry out measures of abortion, private hospitalization, or anonymous placement of the child.

Unmarried mothers from average socioeconomic levels are often able to provide hospitalization or other limited types of care. Many of the illegitimate children from this group are undoubtedly placed by private physicians. However, it is from this group of average or somewhat below average social levels that private child caring and placing agencies tend to receive infants for placement.

Unmarried mothers from inferior socioeconomic levels are primarily characterized by familiarity with various social agencies. Hospitalization is provided at public expense, and frequently the child becomes legally placeable by court order. It is this group that provides most of the children placed by public agencies.

Evidence of the social inferiority of the families in the Skeels-Skodak studies is plentiful:

"A large number of the families were on relief or lived in financial insecurity so marked that it was considered a major factor in making plans for the child. . . . The number of cases in the immediate family with records of mental defect, of dependency involving institutional care, of arrests and incarcerations was larger than would be found in a sample of the same size from the general population" (12, p. 42).

McNemar's major criticisms of studies of children in foster homes are based on his thesis that the true fathers were above average and the true mothers at least average. Hence, the investigators are expected to find a mean IQ, for 154 children, of 116.0 at two years of age and 111.5 at four years, four months, with no child having an IQ below 80 on either first or second examination.

With regard to the occupational status of the true fathers, the figures of the distribution in Table 2 (12, p. 43) speak for themselves:

Occupational Classification	General Population, Employed Males, 1930*	Skeels-Skodak	
		True Fathers Per Cent	Foster Fathers Per Cent
I. Professional	3.1	1.8	13.6
II. Semiprofessional and managerial	5.2	4.6	22.7
III. Skilled trades	15.0	14.6	24.0
IV. Farmers	15.3	10.9	25.3
V. Semiskilled	30.6	10.9	8.4
VI. Slightly skilled	11.3	10.0	4.6
VII. Day laborers	19.5	47.3	1.3

* From Terman and Merrill (15, p. 14).

We believe the tabulation above indicates clearly the selection *downward* of the true fathers and the selection *upward* of the foster fathers.

Our next consideration is that of the education of the true parents. Certainly, a reported mean grade completed of 10.2 for the fathers and 9.9 for the mothers sounds high. But McNemar overlooks a pertinent discussion in the monograph:

"As nearly as can be determined the grades represent grade completed in school. It should by no means be assumed that this represents actual academic achievement of the individual. Judging from those cases where objective school achievement records were available and from the estimates of persons who knew the mothers well, the actual achievement was probably considerably below the school grade attained. That many of the mothers were probably the more inferior members of their school classes is indicated by the large proportion of those who were over age for grade at time of leaving school" (12, pp. 39-40).

In all cases the reported grade was used as the "grade completed." Many of the histories under "education" gave only the statement "seventh grade," "tenth grade," etc., with no indication as to whether this was the grade *reached or completed*.

In Iowa there is compulsory schooling through age 16 (or through eighth grade), usually without benefit of testing programs, psychologists, clinics, or special classes. The dull tend to get "promoted" grade after grade. Sometimes the procedure is to offer music and manual training for duller children in the eighth, ninth, and tenth grades. Such children are permitted to drop some of the academic subjects or to carry them at lower grade levels. The school bus has made high school education available to children even in remote areas. The majority of the unmarried mothers in this group came, not from the strictly rural areas, but from the villages, small towns, and urban areas. There is social pressure for a child to continue in school beyond the eighth grade, regardless of his level of ability.

Intelligence test results were available for 80 of the mothers. The mean intelligence quotient was 88. The following were the chief reasons for not having tests on 74 of the group of 154 mothers: (a) mothers delivering babies in hospitals with no psychologist in residence; (b) a few admittances to the hospital too late in pregnancy to permit a mental examination.

All the clinical materials indicate that the average IQ of the 80

mothers was close to that for the total sampling of 154, "since both examined and unexamined mothers were similar in educational, occupational, and general social background" (12, p. 41).

The fact that the Stanford-Binet is not entirely adequate for adults has been recognized by the investigators and discussed (8, p. 39). For superior adults the scale does not tap the upper limits of ability. Inasmuch as most of the adults in this sampling were failing in the last few years of the Stanford Revision, it is unlikely that many of them reached the test "ceiling." Nor does mental growth cease below the age of 16. The practice of using 16 years as a maximum divisor was in keeping with the explicit instructions of Terman in *The measurement of intelligence* (14, p. 140). If Terman found that 15 years was preferable, to our knowledge that conclusion was never published.

Comments on the test records indicate that the examiners found the IQ's consistent, in most cases, with the girls' general behavior. The tests were not given under unusual stress on the part of the girl. The decision to give up the baby had been made before the test was given; in no case was this influenced by the test results. The mothers had made their major decision and were leading a life of comfort under excellent medical and social supervision. For persons seeking an example of emotional upset in such cases, we recommend examination at the time of the *first discovery* of the pregnancy!

We are glad to say something "about the expertness (or inexperience) of the examiners." They were all experts. These tests were given by the psychologists in the University of Iowa General Hospital, the University Psychopathic Hospital, the Iowa State Board of Control, and the Des Moines Mental Health Center. All these persons were well trained and experienced.

Studies of unmarried mothers who come into agency care are consistent in finding the mean IQ's to be in the upper 80's. On the basis of results on 344 unmarried mothers in maternity wards in seven hospitals in Minneapolis and St. Paul, Lowe (5) reported that 43.3% of these mothers were below IQ 85. (These mothers were examined four weeks *after* the delivery of the child.) Measurement of a population of 7656 school children in this same study showed that only 16.8% were below 85. Lowe's IQ's for unmarried mothers compare closely with the Skeels-Skodak data. Similar results were reported by Nottingham (7).

It is difficult to see how a group in which 54% were below 90 in IQ, 16% were between 70 and 79, and 14% were below 70 could

be judged "average." Even with a smaller divisor for chronological age, the population is weighted toward inferiority.

With reference to Kuhlmann's standardization on young children, McNemar might have turned the page in *A handbook of mental tests* (4), for on page 14 Kuhlmann, having noted this average IQ of 115 for children 6, 12, and 18 months of age, states: "It is seen that these average intelligence quotients are above 1.00 up to the age of four years. This is due to the 'Baby Contest' children that were examined being brighter than the average children."

The Goodenough findings (3) are based on her testing of a sampling of 495 children representing "as closely as possible" the occupational categories found in Minneapolis in 1920. A comparison of this Minneapolis population (Appendix A of Goodenough's "Kuhlmann-Binet Tests") with the distribution of occupations reported by Terman and Merrill (15) indicates a definite selection upward in the Minneapolis group. The Chi-square test shows the difference to be statistically significant.

McNemar states incorrectly that in the Iowa studies the average for the two-year-old children was 108. The only reference to two-year-old children is in Skodak's study (12), Table 5, page 57, in which a mean IQ of 114 is given. The foster children do exceed the averages for unselected children. They exceed the mean predicted from their true parents, even if the latter were, by some stretch of the imagination, considered "average."

If there are any valid objections to the inclusion of tests given at one year, they can be omitted from the major comparisons without affecting the conclusions. As a matter of fact, the trend for differentiation by foster homes is clearer without the tests given at one year. Evidence is lacking to show that the nature of intelligence at two or three years differs markedly from that at four or five.

McNemar is right in holding that Skodak should not have cited Skeels along with Snygg with respect to the correlations between true mother and foster child. Fortunately, she had already stated (12, p. 28): "A preliminary report by Skeels is concerned with the first examinations of children included in this study."

The low relationships between child and true parents are shown by comparisons on IQ levels (12, pp. 84-91), on education reported (12, pp. 76-78), and occupation (12, pp. 78-84). Graphs and tables make the low relationships clear.

McNemar's statements concerning selective placement merely bring into one paragraph what was said on pages 74, 75, 76, and

elsewhere in the monograph (12). What McNemar glosses over is the amount of uniformity in the family backgrounds of the children and the fact that selection was neither consistent nor invariable.

Selective placement was not overlooked as a contributing factor. It was mentioned whenever it may have influenced the comparisons (12, p. 104 and elsewhere). Various subanalyses were included to rule out this factor (12, pp. 82-84; 87-91). Children whose true fathers came from the seventh occupational classification were selected for comparison with the group as a whole. Of this group, those children who were placed in foster homes of the higher occupations were higher in IQ than the children placed in foster homes of the lower occupations. Similarly, children of feeble-minded mothers placed in homes of the three highest categories followed the same general pattern of development as children of all types of background placed in similar homes; likewise, children placed in the four lower categories followed the same developmental pattern as all children placed in such homes.

For children with a final IQ of 115 or over, the mean home inventory score was 95, and for children with a final IQ of 104 or less the mean home inventory score was 77. It is unlikely that all the relationship in the coefficient of .49 is due to the small amount of selective placement.

Wishful thinkers will indeed find it difficult to reconcile maternal IQ's averaging 66 with filial IQ's averaging 108 (11). Skodak included only those mothers with low IQ's who, *in addition*, showed evidence of inability to manage their affairs with ordinary prudence. The other criteria included residence in institutions for feeble-minded or in county homes on wards for the defective; diagnosis of mental deficiency by psychiatrists or psychologists on other grounds than test scores; or the considered judgment of school administrators, social workers, or employers.

The conclusions are based on the indisputable fact that these mothers were known to be feeble-minded. Data on more cases in another state confirming these results are to be found in a study by George S. Speer (13). As our foster children grow older more data will be available. The fact remains that the mental development of this group of children of feeble-minded mothers differs little from that of the other foster children.

McNemar might have included a criticism, raised elsewhere, that appears to the authors as deserving further evaluation. What percentage of all children placed under six months of age have been

examined? The studies by Skeels (8) and Skodak (12) include only children for whom completion of adoption was requested. The suggestion has been made that perhaps the persons who wished to complete adoption were those with children showing promising mental development.

Data are available on children placed from the Iowa Soldiers' Orphans' Home during a five-year period (1933 to 1937, inclusive), and during a three-year period (1934 to 1936, inclusive) from the Iowa Children's Home Society. During these years, 171 children from the Iowa Soldiers' Orphans' Home were placed in adoptive homes under six months of age and 53 from the Iowa Children's Home Society, making a total of 224 children. Of this number, 204 children, or 91.1%, have been given intelligence tests. The mean IQ on this group is 118.7, with a standard deviation of 14.0. This is comparable to the findings on the sampling reported in the Skodak (12) and Skeels (8) studies. There is every reason to believe that the level of intelligence of the 20 children, or 8.9%, who were not examined differs little from the group examined. Of the 20 children not examined, 10 have been adopted.

When a critic makes serious accusations implying omissions and distortions, it is assumed by readers that the entire study has been carefully evaluated. But McNemar fails to mention material that would be of interest to any reader of the criticism. For instance, he does not mention the consistent superiority of children placed in superior foster homes as compared with children placed in less superior homes, even when equated on the basis of true-mother intelligence or true-father occupation. He ignores the small differentiation among the children when classified by maternal IQ level. He does not consider an entire chapter in the monograph (12, pp. 106-126), in which older preschool-age children, tested before placement, showed marked gains in IQ following residence in average and above-average foster homes.

OTHER STUDIES

We agree with McNemar that there are difficulties entailed in converting IQ's into percentiles. The reader will find that most of our articles contain analyses based directly on IQ change. In the first Wellman article (16), both IQ changes and percentile changes were presented. The reader can thus make his own comparisons.

If the test scores of the non-preschool children were high (18) because of frequency of *prior* examinations, the same trend should be found within our preschool groups. It is not. Preschool children gain over both their first and second years in preschool, but fail to gain over the intervening vacation (2, 16, 17). These divergences are not explained on the basis of frequency of prior examination. Also, in Table IX (18, p. 77) it is shown that preschool children of initial levels as high as the non-preschool children gained on retest (initial mean, 120.0; gain, 7.0 points), while the matched non-preschool children lost 3.9 points.

In regard to the article, "Mental growth from preschool to college" (19), McNemar states that the differences between long- and short-attendance children with respect to initial IQ "are large enough to make one question the reality of the later differences on American Council on Education percentiles." In other words, he holds that the difference of 8.5 points in early IQ in the one division (every child below 120 IQ) accounts for a difference of 40.6 percentiles on the high school examination, and that the difference of 1.5 points in IQ in the other division (every child 120 or above in IQ) accounts for a difference of 7.5 percentiles on the high school examination. Since there was an average interval of 10 years between the IQ determination and the high school examination, it appears that McNemar has more faith in IQ's obtained at the preschool ages than we had been led to expect.

Further, McNemar has failed to note that the comparison in Table 5 (19) is between preschool and non-preschool groups, not between long- and short-attendance groups (as was the case in Table 4). The critical ratio of 1.9 in Table 5 indicates that there were 97 chances in 100 that the preschool children were higher on the high school examination than the non-preschool children of the *same* initial IQ and *same* number of years attendance after preschool.

In all cases where individuals have been selected for illustration, this fact has been explicitly stated. We, too, "feel sure that cases among the University school group could be found which would show decrease." Child No. 8 in the selected curves to which McNemar objects is a case in point. This child decreased from an IQ of 155 to 135.

The data on gains in Merrill-Palmer IQ can be reworked without the correlation term, using Fisher's formula. The amount of distortion introduced by the use of cases can be determined from Tables 10

and 11 of the monograph (20), in which the following changes are shown:

Fall to Spring Changes

72 cases: mean gain, 9.1 points IQ; critical ratio, 3.37

42 children with first test in fall and second test in spring:
mean gain, 10.7 points IQ; critical ratio, 3.35

Spring to Fall Changes

46 cases: mean gain, 3.7 points IQ; critical ratio, 1.09

20 children with first test in spring and second test in fall:
mean gain, 5.2 points IQ; critical ratio, .97

Reworking the significance of changes for the individuals, by Fisher's formula, we obtain the following:

42 children, fall to spring: $t = 4.74$ (P is less than .01)

20 children, spring to fall: $t = 2.16$ (P is between .02 and .05)

The difference in change between these two groups is 5.5 points;
 $t = 1.49$. P is between .1 and .2.

In the article by Coffey and Wellman (2), on "The role of cultural status in intelligence changes of preschool children," fall to spring changes in Binet IQ for 417 children were given as follows:

Occupational Group	Children	Mean Change	S.D. of Dist. of Changes
I	278	7.1	11.6
II to IV	139	6.1	11.3

Fall to spring changes are also given for 592 cases, utilizing retests on the above children, as follows:

Occupational Group	Cases	Mean Change	S.D. of Dist. of Changes
I	412	6.0	11.3
II to IV	180	5.7	10.6

Spring to fall changes, however, were negligible. They were as follows:

Occupational Group	Cases	Mean Change	S.D. of Dist. of Changes
I	226	-.62	12.5
II to IV	91	.30	10.6

These differences between fall-spring (preschool attendance) and spring-fall (nonattendance) changes in Binet IQ appear important.

We are unable to find in Terman's instructions for utilization of the Stanford Revision a table indicating the correction to be applied to IQ's of children at different chronological age levels. McNemar's statement indicates a faulty standardization of the test. If this were

known, and generally accepted by the authors of the Stanford Revision, it should have gone out as a supplement to the manual.

If it is true that children are expected to drop in IQ with age, as they go from ages 5 and 7 to ages 12 and 14 (irrespective of the ceiling of the test), then the correction factor to be applied in the Wellman data would indicate gains at the elementary school level even more substantial than those reported.

In any case, on the basis of faulty standardization or an unknown correction factor, it is difficult to account for the wide *divergence* between children in underprivileged homes (Skeels and Fillmore, 9) and the children in our University elementary schools (Wellman, 18).

IQ CHANGES AND REGRESSION

Wellman's discussion of regression and relation to changes on the Merrill-Palmer test was under the heading, "Explanation of Gains" (20, pp. 52-54). That regression could account for the gains was examined as an hypothesis. But it was held in this monograph that regression did not account for the difference between winter and summer gains. It was stated:

"The hypothesis which seems best to satisfy the situation is that a portion of the changes over the winter were direct responses to environmental stimulation. . . . The difference between winter and summer gains may be taken as representing the amount of increase in real ability contributed by the preschool experiences" (20, p. 53).

It is obvious that the more reliable the measuring instrument, the less the regression effect. If retesting yielded the same scores as the original testing, there would be a perfect correlation and no regression phenomena. In establishing the reliability of a test, however, the psychologist runs into practical difficulties. Differences in range of talent and testing interval cannot be ignored. Certainly, an interval long enough to permit real changes in ability, due to environmental conditions, is misleading. For this reason we do not consider as satisfactory reliability coefficients obtained on retests of preschool children at six-month intervals.

Much has been said about the low reliability of tests at the preschool ages. Actually, retest coefficients obtained at the preschool ages fall within the range of reliability at older ages. A retesting at a week's interval on 44 preschool children on the Merrill-Palmer scale yielded an *r* of .90 (20). Goodenough's (3) retesting of preschool children on the Kuhlmann-Binet at a six-weeks interval yielded .82.

Nemzek (6), in 1933, obtained from the literature a median correlation of .83 for repeated testings on the Stanford-Binet for children of all ages. It may be that mental testing at the younger ages is somewhat less reliable than at the older ages, but differences in reliability could not account for the finding of substantial divergences associated with matched groups or different environmental impacts.

When children were retested on the Merrill-Palmer scale at an interval of one week, *all* levels gained, and there was a positive correlation between initial status and gain. It was pointed out in the monograph (20) that these changes did not fit the expected regression trend. How much allowance should we have made for regression in that group? The changes made over six-month periods were inversely related to initial status.

The following points in regard to regression are pertinent to our studies:

(1) Regression does not account for mean gains (such as those made by Iowa City preschool children who started superior and were still higher on retest).

(2) Regression does not account for differences in mean gains of the same children over the winter and summer months (such as found on both Binet and Merrill-Palmer tests).

(3) Regression does not account for differences in trend in two initially equated groups (such as the orphanage preschool and non-preschool groups).

(4) Regression does not account for the high IQ's of the illegitimate foster children. These children did not "regress" toward the mean of their mothers, nor even toward the mean of an unselected population.

* * * * *

With the principle involved in McNemar's last sentence we agree heartily. For scientist, educator, or critic, we feel that *noblesse oblige* is a good doctrine.

BIBLIOGRAPHY

1. BINET, A. *Les idées modernes sur les enfants*. Paris: Flammarion, 1909.
2. COFFEY, H. S., & WELLMAN, B. L. The role of cultural status in intelligence changes of preschool children. *J. exp. Educ.*, 1936-1937, 5, 191-202.
3. GOODENOUGH, F. L. The Kuhlmann-Binet tests for children of preschool age: a critical study and evaluation. *Univ. Minn. Inst. Child Welf. Monogr. Ser.*, 1928, No. 2.
4. KUHLMANN, F. A handbook of mental tests: a further revision and extension of the Binet-Simon scale. Baltimore: Warwick & York, 1922.
5. LOWE, C. The intelligence and social background of the unmarried mother. *Ment. Hyg., N. Y.*, 1927, 11, 783-794.
6. NEMZEK, C. L. The constancy of the I.Q. *Psychol. Bull.*, 1933, 30, 143-168.

McNEMAR'S CRITICAL EXAMINATION OF IOWA STUDIES 111

7. NOTTINGHAM, R. D. A psychological study of 40 unmarried mothers. *Genet. Psychol. Monogr.*, 1937, 19, No. 2, 155-228.
8. SKEELS, H. M. Mental development of children in foster homes. *J. consult. Psychol.*, 1938, 2, No. 2, 33-43.
9. SKEELS, H. M., & FILLMORE, E. A. The mental development of children from underprivileged homes. *J. genet. Psychol.*, 1937, 50, 427-439.
10. SKEELS, H. M., UPDEGRAFF, R., WELLMAN, B. L., & WILLIAMS, H. M. A study of environmental stimulation: an orphanage preschool project. *Univ. Ia Stud. Child Welf.*, 1938, 15, No. 4.
11. SKODAK, M. The mental development of adopted children whose true mothers are feeble-minded. *Child Develpm.*, 1938, 9, 303-308.
12. SKODAK, M. Children in foster homes: a study of mental development. *Univ. Ia Stud. Child Welf.*, 1939, 16, No. 1.
13. SPEER, G. S. The mental development of children of feeble-minded and normal mothers. In *Yearbook of the National Society for the Study of Education. Intelligence: Its Nature and Nurture. Part II. Original Studies and Experiments*. Bloomington, Ill.: Public School Publishing Co., 1940. (In press.)
14. TERMAN, L. M. The measurement of intelligence: an explanation of and a complete guide for the use of the Stanford Revision and extension of the Binet-Simon intelligence scale. Boston: Houghton Mifflin, c. 1916.
15. TERMAN, L. M., & MERRILL, M. A. Measuring intelligence: a guide to the administration of the new revised Stanford-Binet tests of intelligence. Boston: Houghton Mifflin, 1937.
16. WELLMAN, B. L. Some new bases for interpretation of the IQ. *J. genet. Psychol.*, 1932, 41, 116-126.
17. WELLMAN, B. L. The effect of preschool attendance upon the IQ. *J. exp. Educ.*, 1932-1933, 1, 48-69.
18. WELLMAN, B. L. Growth in intelligence under differing school environments. *J. exp. Educ.*, 1934-1935, 3, 59-83.
19. WELLMAN, B. L. Mental growth from preschool to college. *J. exp. Educ.*, 1937-1938, 6, 127-138.
20. WELLMAN, B. L. The intelligence of preschool children as measured by the Merrill-Palmer scale of performance tests. *Univ. Ia Stud. Child Welf.*, 1938, 15, No. 3.

BOOK REVIEWS

WOODWORTH, R. S. Psychological issues: selected papers of Robert S. Woodworth, with a bibliography of his writings. New York: Columbia Univ. Press, 1939. Pp. x+421.

"On October 17th, 1939, Professor Robert Sessions Woodworth achieved three-score and ten . . . His colleagues in the department with which he has so long been identified wished to celebrate this milestone. They decided, however, to do this without distracting him unduly from the scholarly activities that characteristically preoccupied him. Without his knowledge or consent, and with the gracious encouragement of Columbia University, they have, through a small executive committee, prepared this commemorative volume."

Thus commences the Foreword to this collection of psychological articles, which is not only a fitting tribute to one of the most influential scientists of this century in the field of psychology, but at the same time a valuable addition to the psychologist's library.

The complete bibliography of Professor Woodworth's works, which is included in this volume, occupies eleven and one-half pages. Obviously, all of the articles in this list could not be contained in one volume. The members of the committee have used excellent judgment. They have included the most important and interesting treatises and have wisely refrained from revision or abbreviation.

The book begins with the autobiography of Professor Woodworth which appeared in *A history of psychology in autobiography*, and the following twenty-four papers are divided into five sections according to subject matter: systematic problems, abnormal psychology, differential psychology, motor phenomena, educational psychology. It seems proper that the title of the last paper included in this collection of the author of the leading textbooks in the general and in the experimental field should be "The teaching of psychology."

Papers have been selected from the various periods of Professor Woodworth's activity, ranging from his first paper, "Note on the rapidity of dreams" in 1897, to "Situation-and-goal set" in 1937. This list of titles is evidence of Professor Woodworth's widespread interests. One is also impressed with the keenness with which he sensed the important problems throughout the period of rapid development in which he has been active. His contemporaries will relive in memory many of the exciting times of the past decade. His younger colleagues will learn of the important battles that have been fought—the imageless thought controversy, the nature of voluntary movement, the consciousness of relation—to name but a few of them.

The book is invaluable as a cross-section of Professor Woodworth's very sane and scholarly contribution to the science of psychology.

H. S. LANGFELD.

Princeton University.

COLE, L. E. *General psychology*. New York: McGraw-Hill, 1939.
Pp. xii+688.

In the Preface to this volume, the author, with disarming frankness, warns the reader to "beware" of a "behavioristic bias, and of a preference for physiological and mechanistic descriptions of thought and action." He further states that "like many others of the behavioristic persuasion" he possesses "a strong environmentalist bias." Despite this rather formidable front, which will no doubt frighten away some readers, the book does not represent a particularly extreme or radical viewpoint. The appraisal of the reviewer is that it is a good deal more eclectic than its author would have us believe and gives a fair and well-balanced treatment of its subject matter. The "behavioristic bias" of the author simmers down to not much more than an expression of the permanent dent which behaviorism has left in psychology at large. Introspective terms and data are freely used in several chapters, although the author, with characteristic aptness, observes that "the mental contortions involved in trying to have an experience in a normal fashion, and at the same time to scrutinize it critically, like the attempt to study the dark with a flashlight, have not proven very instructive" (p. 342). Adherence to a strict metaphysical behaviorism is definitely weakened by such statements as, ". . . our conscious, rational, deliberative nature seems like a frail and tiny bark swept along on the deeper tides of emotion" (p. 251). The author's position, moreover, does not prevent him from giving such topics as hypnosis, Freudianism, Gestalt psychology, and eidetic imagery treatments which are unusually complete for an introductory text.

The book is exceptionally well written and organized, and has an easy and interesting style. It is, at the same time, scholarly and up-to-date in its handling of material. The author has not followed the practice of some recent textbook writers of breaking up his material into short assignments of ten or fifteen pages each. Instead, he has included only thirteen chapters which average over fifty pages in length. These are arranged in the following topical order: "Animism and Brain Psychology"; "The Central Nervous System and Behavior"; "Receptors"; "The Effectors"; "The Problem of Development"; "Emotion"; "Motivation"; "Learning"; "Perception"; "Thinking"; "Reasoning"; "Intelligence"; and "Personality." The book on the whole is somewhat longer than the usual introductory text, since it contains 688 pages, with index. There are over 100 illustrations, many of which are reproductions from photographs.

It represents clearly the "scientific" or experimental, rather than the "practical" or popularized, approach, and is rich in references which are given in the form of footnotes. In some cases the author devotes several pages to the discussion of a single experiment which he considers particularly significant. The nervous system, effectors, and receptors receive especially detailed and complete treatment. Considered altogether, the selection of material is well rounded and complete. The only important omission from which the book seems to suffer is its apparent aversion for anything related to mathematical concepts in psychology. Thus, the Weber-Fechner Law, psychophysics or its derivatives, the normal proba-

bility curve, correlation coefficient, and standard deviation receive little or no mention.

The author, on the other hand, makes more of controversial issues than is customary in an introductory text. Many of the chapters are introduced by showing their controversial backgrounds. In this way he has, of course, made use of the historical approach; but he has also presented, at the very outset, material which some writers would have left till the end of their handling of the topics in question. Examples of this arrangement are to be found in the fact that a discussion of the nature of the mind and the soul—going back to Hippocrates, Aristotle, and Galen—gives the student a general introduction to psychology as a whole and to the nervous system in particular. The chapter on "The Problem of Development" begins with the instinct controversy. The chapter on "Emotion" is introduced by the James-Lange controversy. The subject of "Attention" starts with a critical evaluation of its mentalistic limitations; and "Learning" is introduced with the problem of whether or not it should be defined in terms of the nervous system. The discussions are very much to the point and should prove of interest to the advanced student. One wonders, however, just how much the neophyte will get out of them.

Cole's *General psychology* is not an "easy" book, but should prove stimulating as well as teachable for the intelligent student. It should satisfy two definite sorts of needs in college teaching. First, it should meet the demands for a solid, conservative, scientific text for short introductory courses in which some selection of student personnel can be exercised. Second, it is adapted to longer or more concentrated introductory courses where no student selection is possible. The point of view is modern and refreshing. The author has pressed his behavioristic threat far enough to drive out undesirable spooks. At the same time he has been willing to sacrifice his consistency of approach enough to include important topics which could not otherwise be treated.

W. N. KELLOGG.

Indiana University.

SYMONDS, P. M. *The psychology of parent-child relationships.* New York: Appleton-Century, 1939. Pp. xiv+228.

This book, by an author who is well known as an authority in the field of personality and measurement, should have a wide appeal to persons interested in child development and in the formation and structure of the personality. The concept of emotional security in parent-child relationships is keynoted in examination of psychoanalytic and experimental literature, in two experimental studies carried on by the author and here reported in detail, and in the author's theoretical analysis of the general problem. Further application of the findings is made to teacher-child and counselor-client relationships, both of which are regarded as basically similar to the parent-child relationship.

Dr. Symonds has two original and worthy aims in his plan for this book. He is seeking, first, to correlate mentalistic concepts derived from psychoanalysis with bits of behavior obtained from relatively more super-

ficial observation and, secondly, to incorporate in one volume a scientific experimental program with its findings, together with practical discussion of the merits of one or another technique of handling children so as to produce socially desirable results. The first aim is carried out in an interesting way and gives results which are at least provocative, if somewhat difficult of interpretation in any but extremely broad terms. For example, in a study of the mental life of children whose parents were either clearly dominating or clearly dominated by them, one of the questions asked was: "Does the child resent parental authority? School authority?" The answer, made on the basis of interview material gathered by teachers, counselors, and school principals, is definite for these groups of children. The dominated children do not resent authority and are not rebellious toward it. Rebellion and resentment occur much more frequently among the dominating children. This and other data give a picture of the dominated child as apathetic, humble, and shy. The reviewer doubts whether the psychoanalysts who first brought out the significance of parental dominance would accept this picture as more than surface behavior adopted for the benefit of the interviewer who, as Dr. Symonds brings out clearly in later sections of the book, stands in the role of parent to the child. The children of dominating parents have learned by experience to anticipate punishment for displays of temper and rebellion; thus, in a situation in which they are not strongly frustrated (interview with a sympathetic investigator) they react with the retiring behavior here given as typical. It is yet to be proved that this reaction is characteristic for more strongly motivated behavior or for more than the surface aspects of their mental life. Clinical evidence suggests that it does not work out in this way for the majority of dominated children.

The second aim of the work, the application of experimental findings to practical problems of management of the parent-child and parent (substitute)-child relationship, appears to give the author some difficulty in appeasement of his scientific conscience. The Hitler and Stalin of this piece are, of course, the twin bugaboos of causation and value. The impossibility of finding causes by this sort of experimental work, and, on the other hand, the necessity for causal relationships to appear fairly clearly for applications of knowledge to be useful, makes for an author's dilemma which is exemplified by the following statement: ". . . the chances are far higher than 99 in 100 that accepted children show more socially acceptable behavior . . . than rejected children. Of course this is not at all the same thing as saying that behavior which is socially disapproved is caused by parental rejection and our data by no means prove this. It is presumably true, however, that parental acceptance or rejection is a primary factor in a person's life and hence in personality formation, and it is reasonable to suppose that the figures presented do indicate causal relationship to a considerable degree" (p. 73). In the same way it is necessary, in writing of parental attitudes which will produce optimally stable children, to set up a series of value judgments in order to picture the adjusted child who is the goal of the procedure. Such judgments are, of course, made every day by the therapist and educator. The difficulty arises in trying to make them *on the basis of*

experimental findings. A connection of this sort (which it must in fairness be said is not explicitly made by Dr. Symonds) is not good practice, not so much for its lack of scientific rigor as for its effect on the various kinds of lay people who avidly read books of this sort.

These minor objections in no way detract from the value of the book for psychologists and educators. There is a wealth of interesting material, gathered from both experimental and theoretical approaches, which represents a comprehensive summary of existing knowledge and opinion on the subject of parent-child relationships as well as a forward-looking contribution to the field of mental hygiene.

PAULINE SNEEDEN SEARS.

Yale School of Medicine.

MONROE, P. (Ed.) Conference on examinations (under the auspices of the Carnegie Corporation, the Carnegie Foundation, the International Institute of Teachers College, Columbia University, at the Hotel Royal, Dinard, France, September 16th to 19th, 1938). New York: Bureau of Publications, Teachers College, Columbia Univ., 1939. Pp. xiii+330.

This book is a record of the fifth session of the International Conference on Examinations held at Dinard, France, during September, 1938. It includes reports and general discussions by committee members representing France, England, Scotland, Sweden, Norway, Finland, and the United States.

The French committee reviews its progress in investigating four problems: (1) the marking of papers at the written examination of the baccalauréat; (2) the interdependence of intellectual aptitudes as revealed by the marks of students at the written examinations; (3) the relations between the École Normale Supérieure and the examination system; and (4) the selection and orientation of children who are entering the first form—the lowest form of the secondary school. The major studies of the Scottish committee are concerned with the intelligence of Scottish children, the prognostic value of university entrance examinations, qualifying examinations, and a follow-up study of individual mental testing. The Swedish group reports it is studying entrance examinations in secondary schools, marking or levels of marking in elementary schools, the influence of entrance examinations on previous teaching, and the matriculation examination. The Norwegians indicate that their great problem arises from the fact that democracy allows so many pupils to pass a secondary school. The Finnish representatives relate their concern with a general review of developments with regard to examinations, with matriculation examinations, and with entrance examinations in secondary schools.

The most extensive studies of the marking of essay examinations are reported by the English committee. They recognize that in carrying on these investigations they are repeating much of the work done elsewhere, especially in America, and state frankly that a very large part of their task "has really been to educate the English public in certain directions." The major study reported by this group was carried on in twenty-four

schools where the pupils of about sixteen years of age wrote six essays, four objective tests, and an intelligence test. Some of the essays were written on undirected subjects according to the general formula: "Write anything about the subject for anybody." Other essays were written on directed subjects and for specific readers. Each essay was marked for not less than seven characteristics: sense, spelling, punctuation, grammar, vocabulary, sentence structure, and general impression. The conclusions indicate (1) no differences in results of marking the directed and undirected essays; (2) that sense and general impression give about the same percentage of marks; and (3) that it would be unjust to estimate the ability of any candidate by a single, or even by two, essays.

The United States is represented in the volume through a summary of the well-known Pennsylvania Study by W. S. Learned, a review of the work of the Coöperative Test Service by Ben D. Wood, a discussion on the purity of tasks in examinations and on the importance of scaled tests by E. L. Thorndike, and, finally, a brief summary of the conference by I. L. Kandel.

The report as a whole recalls the early studies in the United States on the reliability of essay examinations and emphasizes similar conclusions. Following the insistence upon the need for subjective examinations by a French representative, Kandel summarizes the present status of examinations as follows: "I do not think anybody who has studied the problem seriously would say that the question is either/or; the question is, how can we today provide the best education for each individual, and to what extent can we use whatever measures we may discover, not excluding subjective."

Like most reports of conferences, this volume includes much tedious and inconsequential material. Its weaknesses are those of the *Congressional Record*. The chief value of the book lies in its re-emphasis upon the common inadequacies of current examinations in all countries and its stress upon the need for intensive research to improve such measures.

ALVIN C. EURICH.

Stanford University.

COLETTE, E., et al. *Le mystère animal. ("Présences.")* Paris: Plon, 1939. Pp. ix+301.

In the Preface to this collection of ten essays, the belief is expressed that "although the book brings forward questions perhaps less burning than those of communism or liberty," the study of animals may well advance our knowledge "of this true 'humanism' which is the aim of the *Présences* . . . We study animals to know them, without doubt; but is it not our ulterior design to detect, through them, some of the essential secrets of man? There is a mystery of animals which throws light on the mystery of man." Although the present volume does not penetrate deeply into either of these mysteries, it does point out directions which the search may take and presents a miscellany of facts concerning animals and their place in this man-dominated world.

An imaginative introduction by Colette, writer of animal stories, is followed by Part I: five chapters, by as many different authors, includ-

ing facts and anecdotes about animal behavior, interpretations and reflections, and considerations pertaining to the usefulness of animals and man's responsibilities towards them. By far the most substantial of these chapters is by Daniel-Rops, editor of the *Présences* series. In speculating about the motives for the modern, materially unselfish interest in animals, the author finds a significant residue after taking account of the purely scientific aims of "pushing back the limits of objective knowledge" and of "fixing the exact limits that separate the most intelligent primate from the most primitive pygmy." He asks whether those people who "love animals and are glad to see them live, but who yet have no desire to probe the secrets of life," are not perhaps driven by an obscure nostalgia for the original paradise where man and animal lived in peace and harmony. Daniel-Rops concludes that "the time of the earthly paradise is no more and all our love for animals would not know how to find the road again."

The five chapters of Part II are written by specialists, the first four dealing with comparative neurology, psychology, pathology, and sociology, respectively, and the final chapter, by Abbé Jean Plaquevent, with the problem of the soul. These contributions may be valued more for their stimulating observations and provocative ideas than as comprehensive and completely reliable summaries of the several fields treated. Claparède, in his chapter, "From Animal Intelligence to Human Intelligence," concludes that the superiority of human intelligence is intimately related to freedom from, and awareness of, the environment, which accompany the development of language.

Le mystère animal is a curious blend of science, sentimentality, politics, philosophy, and religion. The dominant note is not easy to determine. Although the volume obviously is intended primarily for the intelligent lay reader, the scientist may find entertainment and stimulation in its unusual perspectives.

HENRY W. NISSEN.

Yale Laboratories of Primate Biology.

GILLILAND, A. R., & CLARK, E. L. *Psychology of individual differences.* New York: Prentice-Hall, 1939. Pp. xvi+535.

These authors announce that their purpose is "to collect in one volume the more important facts and conclusions in the broad field of individual differences." A further statement asserts that "complete changes in the interpretation of the work of others which are sometimes attempted are rarely valid, and are not attempted in this book." These remarks, it turns out, serve fairly well to define the scope and to suggest the general character of the work. It is primarily a comprehensive "collection" of facts, rather than a critical appraisal of the scientific status of the facts collected.

It seems to the reviewer that one consequence of the attitude of methodological neutrality adopted by the authors is failure to achieve a logical organization of the subject matter. This outcome is forecast in the first chapter, in a discussion of the problem of classification, in which it is proposed that "the ways in which people differ can be classified as physical, mental, educational, and personality differences." While it is

perhaps possible to define these terms so as to designate somewhat diverse dimensions of human variability, it would seem that the authors scarcely have accomplished this end. Within the category of "physical differences," for example, are to be found such diverse phenomena as height, health, reaction time, and skill in playing golf. (The special chapter on "Differences in Physical Characteristics" is, however, restricted mainly to anthropometric measurements.)

The choice and arrangement of chapters illustrates further the disposition on the part of the authors to assort the "facts" under conventional categories, without much regard for systematic considerations. The following sequential order of chapter headings will perhaps illustrate this tendency: (1) "The Problem of Individual Differences"; (2) "The Causes of Differences"; (3) "The Measurement of Differences" (statistical section); (4) "Differences in Physical Characteristics"; (5) "Sex Differences"; (6) "Race Differences"; (7) "The Influence of Near Ancestry"; (8) "Differences in Intelligence"; (9) "Types of Extreme Deviations"; (10) "Individual Differences in Personality"; (11) "Individual Differences in Learning and Teaching"; (12) "Applications of Individual Differences in Business and Industry"; (13) "The Significance of Individual Differences." It would seem, for example, that Chapter 2, on "The Causes of Differences," might well have been related to Chapter 7, "The Influence of Near Ancestry." But these two chapters stand apart as independent units. And the book as a whole could have been organized into three general divisions: (1) the scientific study of individual variability; (2) group differences; (3) the significance of individual and group differences for education, industry, and social organization. But the authors preferred to use a somewhat heterogeneous collection of conventional categories and to impose upon them a minimum degree of methodological "regimentation." The general result is that the several chapters stand somewhat in isolation, one from another, and the book accordingly might be regarded as an elementary "source book" rather than as a systematic textbook.

It should be noted, however, that this relative independence of the individual chapters has certain advantages. For one thing, it will permit the use of the book as a convenient reference work in introductory courses in psychology which attempt to include this field. The chapter on "The Influence of Near Ancestry," for example, is a very good review of the major studies in this field, as is the chapter on "Sex Differences." And throughout the book there is a wealth of original data instead of generalized paraphrasing. Among other good features, the following deserve special mention: the judicious handling of the "nature-nurture" problem; the bountiful supply of tables and charts; the well-chosen lists of questions and exercises at the end of each chapter; the bibliographical materials (both in footnotes and at the ends of chapters).

The book will undoubtedly prove to be a valuable addition to the instructional resources of one of the few branches of psychology which has not yet been inundated with textbooks. Its value as a textbook will undoubtedly be considerably enhanced, for many psychologists, by the absence of a doctrinaire tone. And for such psychologists this virtue

will perhaps more than compensate for any deficiencies in organization or in methodological evaluation of materials.

LYLE H. LANIER.

Vassar College.

ANKLES, T. M. *A study of jealousy as differentiated from envy.* Boston: Bruce Humphries, 1939. Pp. 109.

The purpose of *A study of jealousy* is to analyze this emotion "with a view to finding the exact components, the way they work, the external factors which render this behavior recognizable, and the internal mechanism." Ankles attempts to show that jealousy, a complex emotion, has the same components as an organic drive. (1) The incentive which evokes the emotion is *fear of loss* of someone in whom the person feels some right of possession. (2) The drive is *sadism-masochism*. This is innate, is coördinate with, and, anthropologically, a derivative of, the sex drive. (3) *Self-feeling* determines whether the sadistic or masochistic tendency is dominant. If self-feeling is high, masochism is dominant; if low, sadism is dominant. (4) The *affective tone* of the satisfaction of one or more of the above components is always pleasant. The data are the "introspective" reports of thirty subjects met at the tea table in public restaurants. Engaging the subject in conversation, Ankles questioned him concerning his feelings of jealousy, making notes after he left.

The reviewer's chief criticisms are these: (1) Trained introspectionists were not used. Only ten subjects could give a satisfactory definition of jealousy. A study so dependent upon introspection requires the use of more carefully controlled reports than can be secured from tea-time acquaintances. (2) The theory was not derived, as purported, from the data. There are several instances which indicate that Ankles forced the data to fit a theory he had at the outset of the inquiry. Pleasure sometimes was not mentioned as a felt component of jealousy. Ankles then would explain his theory and ask if the subject now saw how pleasure was involved. In case this was denied categorically Ankles explains that the pleasure is unconscious, having been repressed by an unconscious set of ideas opposed to the expression of pleasure as a conscious component of jealousy. (3) The data exhibit obvious discrepancies. Although only fourteen subjects are "introverts," twenty-eight report self-feeling, characteristic supposedly only of introverts. (4) The validation of the theory is quite inadequate. Questionnaire returns from ten subjects ("the number being carefully restricted") supply the validating data. To one schooled in the philosophy of statistical method comment hardly seems worth while.

In spite of these criticisms the reviewer feels that the theory should be valuable as a working hypothesis in research. Ankles has tried to formulate an objective theory. He distinguishes it (verbally) from envy and hate. He shows the theory to be consistent with certain of those advanced by Freud, Briffault, Stekel, Shand, and others who have studied the anthropology and evolution of sex.

LEONARD W. FERGUSON.

University of Connecticut.

BOOKS RECEIVED

BARTHEL, E. *Der Mensch und die ewigen Hintergründe: Religionsphilosophie, Metaphysik der Zeit und ethische Zielbestimmung.* München: Ernst Reinhart, München 13, Isabellastrasse 11, 1939. Pp. 70.

EDDINGTON, SIR A. *The philosophy of physical science.* New York: Macmillan; Cambridge, England: University Press, 1939. Pp. ix+230.

KLEIN, M. *Von der All-Einheit im Ich: eine paradigmatische Philosophie.* München: Ernst Reinhart, München 13, Isabellastrasse 11, 1939. Pp. 240.

KRUEGER, F. Otto Klemm und das psychologische Institut der Universität Leipzig: deutsche Seelenforschung in den letzten drei Jahrzehnten. Leipzig: Johann Ambrosius Barth, Salomonstrasse 18b, 1939. Pp. 94.

LEDERER, R. K., & REDFIELD, J. *Studies in infant behavior V.* Univ. Ia Stud. Child Welf., Vol. XVI, No. 2. Iowa City: University, 1939. Pp. 157.

REED, H. B. *Psychology and teaching of secondary-school subjects.* New York: Prentice-Hall, 1939. Pp. xviii+684.

WALSH, G. *Sing your way to better speech: a jingle sequence for the improvement of articulation and rhythm in speaking.* New York: Dutton, 1939. Pp. 209.

WINKLER, J. K., & BROMBERG, W. *Mind explorers.* New York: Reynal & Hitchcock, 1939. Pp. 378.

WOODWORTH, R. S. *Psychological issues: selected papers of Robert S. Woodworth, with a bibliography of his writings.* New York: Columbia Univ. Press, 1939. Pp. x+421.

NOTES AND NEWS

DR. ANNE ANASTASI has been appointed chairman of the department of psychology at Queens College, New York City.

DR. CHRISTIAN A. RUCKMICK has joined the firm of C. H. Stoelting Company as secretary of the firm and general sales manager.

DR. ROBERT MACDOUGALL, professor emeritus of psychology at New York University, died on October 31. He was 73 years of age.

DR. FRANK ANGELL, professor emeritus of psychology at Stanford University, California, died on November 2, at the age of 82 years.

MORE than 60 professional psychiatrists and psychologists from eastern Massachusetts were guests, on the evening of November 13, of three members of the psychiatry department of the Tufts College Medical School and the nine members of the American Psychological Association who are on the Tufts faculty, at a meeting at the home of Leonard Carmichael, President of Tufts College and President of the American Psychological Association.

Among those who spoke briefly on the relationship between psychiatry and psychology were: Douglas A. Thom, professor of psychiatry, Tufts Medical School; Abraham Myerson, professor of neurology, Tufts Medical School; A. Warren Stearns, professor of psychiatry and dean of the Tufts Medical School; C. Macie Campbell, professor of psychiatry, Harvard Medical School; Truman Lee Kelley, professor, Graduate School of Education at Harvard; Edwin G. Boring, professor of psychology, Harvard; Gordon Allport, professor of psychology, Harvard; Ross A. McFarland, of the Fatigue Laboratory, Harvard Business School; Edna Heidbreder, professor of psychology, Wellesley College; David Shakow, clinical psychologist, Worcester State Hospital; Vernon Jones, head of the psychology department, Clark University; Hudson Hoagland, head of the biology department, Clark University; and E. Stanley Abbot, psychiatrist and psychologist, of Boston.

THE Washington-Baltimore branch of the American Psychological Association held its fall meeting on November 2, 1939, at the Johns Hopkins University. The following program was presented:

ELIZABETH RUTHERFORD: "The Remedial Reading Program at Goucher College."

LOUISE SLOAN: "Recent Developments in the Dark Adaptation Test of Vitamin A Deficiency."

J. M. STEPHENS: "Alternative Explanations of the Results of Experiments on Symbolic Punishment."

THE Seventeenth Annual Meeting of the American Orthopsychiatric Association, an organization for the study and treatment of behavior and its disorders, will be held at the Hotel Statler, Boston, Massachusetts, on February 22, 23, and 24. Dr. Norvelle C. LaMar, 149 East 73rd Street, New York City, is the secretary of the Association.

THE report on Project No. 2 of the Inter-Society Color Council, published in the September issue of the *Journal of Research of the National Bureau of Standards*, is now available in reprint form. Copies may be purchased from the Superintendent of Documents, Government Printing Office, Washington, D. C., for 10 cents each. The problem dealt with in this report caused the formation of the Inter-Society Color Council, and was "to find a means of designating colors in the U. S. Pharmacopoeia, the National Formulary, and pharmaceutical literature; such designation to be sufficiently broad to be appreciated and usable by science, art, and industry, and sufficiently commonplace to be understood at least in a general way by the whole public." For the past seven years committees of the Council have been working on a solution. In June, 1939, a solution reported by Mr. Kenneth L. Kelly, of the American Pharmaceutical Association, and Dr. Deane B. Judd, of the National Bureau of Standards, was accepted by letter ballot of the voting delegates of the Inter-Society Color Council. The Council has recommended the method for designating the colors of drugs and chemicals, but final action on recommendation for general use will not be taken until there has been opportunity to study its application to color work covered by the interests of member bodies. Comments on the applicability of this method should be sent to Deane B. Judd, National Bureau of Standards, Washington, D. C.

THE Ninth Annual Meeting of the Inter-Society Color Council is to be held jointly with the Optical Society of America and the American Physical Society, February 21-24, in New York City. The technical session on "Spectrophotometry in the Pulp and Paper Industry," to be held on the afternoon of February 21 at the Roosevelt Hotel, will be conducted by John L. Parsons, of the Hammermill Paper Company, and will include an introductory dialogue and discussions relative to the Survey and Use of Instruments. In the evening a popular session will be held in the auditorium of the Electrical and Gas Association of New York, 480 Lexington Avenue. Requests for tickets for this session, accompanied by a stamped, self-addressed envelope, should be addressed to Inter-Society Color Council, P. O. Box 155, Benjamin Franklin Station, Washington, D. C. On February 22 there will be a morning discussion session, under the chairmanship of Deane B. Judd, which will be continued in luncheon groups at the Roosevelt Hotel, followed by a business meeting in the afternoon. Joint meetings with the Optical and American Physical Societies will take place at Columbia University on February 23 and 24. These will include a symposium on "Optical Methods for the Study of Molecular Structure."

slightly quizzical and ready to guess. I am a thousand times
more interested in finding out what she will do than in predicting
what she will do. I am still curious about her and her husband.

She is a woman who has had a good deal of life experience.
She has been married twice, she has had two children, she has
had a good deal of time to think and to reflect. She is a woman
who has had a good deal of time to think and to reflect. She is a
woman who has had a good deal of time to think and to reflect.

She is a woman who has had a good deal of time to think and to reflect.
She is a woman who has had a good deal of time to think and to reflect.

She is a woman who has had a good deal of time to think and to reflect.
She is a woman who has had a good deal of time to think and to reflect.

She is a woman who has had a good deal of time to think and to reflect.
She is a woman who has had a good deal of time to think and to reflect.

She is a woman who has had a good deal of time to think and to reflect.
She is a woman who has had a good deal of time to think and to reflect.

She is a woman who has had a good deal of time to think and to reflect.
She is a woman who has had a good deal of time to think and to reflect.

She is a woman who has had a good deal of time to think and to reflect.
She is a woman who has had a good deal of time to think and to reflect.

She is a woman who has had a good deal of time to think and to reflect.
She is a woman who has had a good deal of time to think and to reflect.

She is a woman who has had a good deal of time to think and to reflect.
She is a woman who has had a good deal of time to think and to reflect.

